

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:01 ; Search time 94.8513 Seconds
(without alignments)
6041.216 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MWSEEEEDGDAETQDSE.....KLKGRKSRSHNNKLIQSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14333	100.0	2781	4	Q9UGI2	Q9UGI2 homo sapien
2	2839.5	19.8	2669	5	Q95V8	Q95V8 drosophila
3	2775.5	19.4	645	11	Q8VDN7	Q8VDN7 mus musculu
4	1557.5	10.9	976	5	Q9W0T0	Q9W0T0 drosophila
5	1368.5	9.5	412	4	Q9H5E0	Q9H5E0 homo sapien
6	871.5	6.1	1711	5	Q45409	Q45409 caenorhabd
7	865.5	6.0	1713	5	Q95Z08	Q95Z08 caenorhabd
8	674.5	4.7	1022	5	Q960Y3	Q960Y3 drosophila
9	672.5	4.7	997	5	Q9W0T1	Q9W0T1 drosophila
10	572.5	4.0	2768	5	Q9VC00	Q9VC00 drosophila
11	572	4.0	17352	5	Q95Y02	Q95Y02 procampar
12	560	3.9	5476	5	Q9N17	Q9N17 drosophila
13	560	3.9	5533	5	Q9VPL2	Q9VPL2 drosophila
14	560	3.9	5533	5	Q9U6C3	Q9U6C3 drosophila
15	560	3.9	5554	5	Q9NHN1	Q9NHN1 drosophila
16	560	3.9	5560	5	Q9VPL1	Q9VPL1 drosophila

17	549.5	3.8	6815	5	Q917U4	Q917U4 drosophila
18	549.5	3.8	16215	5	Q9NFS3	Q9NFS3 drosophila
19	534	3.7	3080	5	Q9VRY3	Q9VRY3 drosophila
20	529.5	3.7	3257	5	Q9V736	Q9V736 drosophila
21	521	3.6	2465	5	Q22463	Q22463 caenorhabd
22	508.5	3.5	2897	5	Q9VTD9	Q9VTD9 drosophila
23	494.5	3.5	5327	5	Q76891	Q76891 drosophila
24	488	3.4	2910	10	Q9FND5	Q9FND5 arabidopsis
25	478.5	3.3	2081	10	Q9LH98	Q9LH98 arabidopsis
26	459	3.2	5170	5	Q17490	Q17490 caenorhabd
27	459	3.2	6994	5	Q17343	Q17343 caenorhabd
28	441.5	3.1	4498	5	Q9W223	Q9W223 drosophila
29	437	3.0	5120	13	Q9PU36	Q9PU36 gallus gall
30	434.5	3.0	2977	5	Q9VAP9	Q9VAP9 drosophila
31	431.5	3.0	2402	2	Q9AER7	Q9AER7 staphylococ
32	430.5	3.0	3261	4	Q9Y556	Q9Y556 homo sapien
33	430.5	3.0	3664	4	Q96T58	Q96T58 homo sapien
34	430	3.0	1795	5	Q76894	Q76894 drosophila
35	430	3.0	2586	5	Q9VTK8	Q9VTK8 drosophila
36	429	3.0	3726	5	Q9VPL1	Q9VPL1 drosophila
37	427	3.0	1483	5	Q8T273	Q8T273 dictyostell
38	423	3.0	2285	5	Q9VP17	Q9VP17 drosophila
39	421.5	2.9	5147	4	Q9Y6V0	Q9Y6V0 homo sapien
40	419.5	2.9	1850	11	Q91YF5	Q91YF5 mus musculu
41	419.5	2.9	2232	5	P91365	P91365 caenorhabd
42	418	2.9	2994	5	Q95ZG5	Q95ZG5 dictyostell
43	418	2.9	4025	4	Q9NR13	Q9NR13 homo sapien
44	415.5	2.9	3484	5	P91257	P91257 caenorhabd
45	413.5	2.9	3507	5	Q23587	Q23587 caenorhabd

ALIGNMENTS

RESULT 1
ID Q9UGI2 PRELIMINARY; PRT; 2781 AA.
AC Q9UGI2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bromodomain PHD finger transcription factor.
GN BPTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20130111; PubMed-10662542;
RA Jones M.H., Hamana N., Shlman M.;
RT "Identification and characterization BPTF, a novel bromodomain
transcription factor";
RL Genomics 63:35-39(2000).
DR EMBL; AB032251; BAA89208.1; -.
DR HSSP; Q92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT dom.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
SQ SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1AC3CFBE CRC64;

Query Match 100.0%; Score 14333; DB 4; Length 2781;
Best local similarity 100.0%; Pred. No. 0;

Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MSEEEEDDGA	ETDODSEDEDEDEDDDDSDYPEEMEDDDDA	SYCTESSFRSH	60				
Db	1	MVEEEDDGA	ETDODSEDEDEDEDDDDSDYPEEMEDDDDA	SYCTESSFRSH	60				
QY	61	YSTPGRRKPRVHR	PRSPILKEKDIPLEPKSSSEDLAMVNEHIMNIAI	YEVLRNGV	120				
Db	61	YSTPGRRKPRVHR	PRSPILKEKDIPLEPKSSSEDLAMVNEHIMNIAI	YEVLRNGV	120				
QY	121	LRSPREFEDPCALY	SOECCTLMAEMHVYLKAVLREEDTSNTTGP	ADLKS	180				
Db	121	LRSPREFEDPCALY	SOECCTLMAEMHVYLKAVLREEDTSNTTGP	ADLKS	180				
QY	181	FIDGMTPVELRY	YCESDKRYHHVLPYOEADY	PYGVENEKIKVLOFLV	QFTTNARE	240			
Db	181	FIDGMTPVELRY	YCESDKRYHHVLPYOEADY	PYGVENEKIKVLOFLV	QFTTNARE	240			
QY	241	ELMSEGVIOYDHC	RYCHKLGDLCCETCSAVYHLECVKPLLE	VEPEDEMOCEVCANHY	300				
Db	241	ELMSEGVIOYDHC	RYCHKLGDLCCETCSAVYHLECVKPLLE	VEPEDEMOCEVCANHY	300				
QY	301	PGYTDCAEAO	QKPKYTRHPITGDSRKRYWFLNRLLI	EEDTENENEKIMYSTKVQ	360				
Db	301	PGYTDCAEAO	QKPKYTRHPITGDSRKRYWFLNRLLI	EEDTENENEKIMYSTKVQ	360				
QY	361	LAELIDCLDKDY	WAEELCKLLEEMREIHRHMOTEDLT	TKARGSNKSF	LAANEELLES	420			
Db	361	LAELIDCLDKDY	WAEELCKLLEEMREIHRHMOTEDLT	TKARGSNKSF	LAANEELLES	420			
QY	421	IRAKKGDIDNV	KSPETEEDKNETENDSKDAENREEFEDOS	LEKSDDKTPDDPEOGK	480				
Db	421	IRAKKGDIDNV	KSPETEEDKNETENDSKDAENREEFEDOS	LEKSDDKTPDDPEOGK	480				
QY	481	SEVGDKRSEKSN	ELSPSPAGGASGSTRITRLRNPDSKL	SOLKSQOYAAAAH	ENKL	540			
Db	481	SEVGDKRSEKSN	ELSPSPAGGASGSTRITRLRNPDSKL	SOLKSQOYAAAAH	ENKL	540			
QY	541	FKGKGEVLVNS	OGELSRLLSTKKEVIMKGINNYFKLG	GGKRYVYNOYSTNS	FALNKH	600			
Db	541	FKGKGEVLVNS	OGELSRLLSTKKEVIMKGINNYFKLG	GGKRYVYNOYSTNS	FALNKH	600			
QY	601	QHRHDHDKRHL	AHKFLCPAGEFKWNGSVHSGKVLIT	SLRLTLTTOLENNI	ESSFLHPN	660			
Db	601	QHRHDHDKRHL	AHKFLCPAGEFKWNGSVHSGKVLIT	SLRLTLTTOLENNI	ESSFLHPN	660			
QY	661	WASHBRANWIK	AYOMCSRPREFALALILECAVYKPVMLP	TWREPLGTRLRHMTS	IERE	720			
Db	661	WASHBRANWIK	AYOMCSRPREFALALILECAVYKPVMLP	TWREPLGTRLRHMTS	IERE	720			
QY	721	KEKVKKKKKQ	EEETMOQATWVYKTPYVAKHQWKOG	EEYRTYGGGMSI	KTHVYRF	780			
Db	721	KEKVKKKKKQ	EEETMOQATWVYKTPYVAKHQWKOG	EEYRTYGGGMSI	KTHVYRF	780			
QY	781	VPLPGNTNNTN	YKSLLEGTKNNMDENNDSDKRC	SBSPKIKIEPDS	EEKVKGSDAAK	840			
Db	781	VPLPGNTNNTN	YKSLLEGTKNNMDENNDSDKRC	SBSPKIKIEPDS	EEKVKGSDAAK	840			
QY	841	GADONENDISK	ITREKDDOYKELLSDSDKPCKEE	PEVNDMMKTS	SHVNCSSQYDVY	900			
Db	841	GADONENDISK	ITREKDDOYKELLSDSDKPCKEE	PEVNDMMKTS	SHVNCSSQYDVY	900			
QY	901	NVSEGFHLRT	SYKKTSSKLDGLERRIKOFL	EEKORLEKTKLEG	IGIKGTSTNSS	960			
Db	901	NVSEGFHLRT	SYKKTSSKLDGLERRIKOFL	EEKORLEKTKLEG	IGIKGTSTNSS	960			
QY	961	KNJSESPVIT	KAEGCOSDSMRQOSPANNNDP	EDLIQCSQSDSSV	LRMSDP	SHTTNK	1020		
Db	961	KNJSESPVIT	KAEGCOSDSMRQOSPANNNDP	EDLIQCSQSDSSV	LRMSDP	SHTTNK	1020		
QY	1021	LYPKDVLVDV	STRSPETKCPKONSTENDLE	EKVSLASGQPT	SKTGNDF	IDDSK	1080		
Db	1021	LYPKDVLVDV	STRSPETKCPKONSTENDLE	EKVSLASGQPT	SKTGNDF	IDDSK	1080		

QY	1081	LASADIGTLIC	KNKRPILQEEPTIVSSSKSALHSSV	PKSTNDRD	ATPLSRAM	DEGKL	1140										
Db	1081	LASADIGTLIC	KNKRPILQEEPTIVSSSKSALHSSV	PKSTNDRD	ATPLSRAM	DEGKL	1140										
QY	1141	GCDSESNSTL	ENSDDVYSDSSSEDMIVONSNESIS	SEQFRTRQ	QDVLEPL	CELVSG	1200										
Db	1141	GCDSESNSTL	ENSDDVYSDSSSEDMIVONSNESIS	SEQFRTRQ	QDVLEPL	CELVSG	1200										
QY	1201	ESTGNCDR	PLPVGTENGKRP	PSQOKLERRPVNKC	QDKLKN	TDPKKNNE	RESKKG	1260									
Db	1201	ESTGNCDR	PLPVGTENGKRP	PSQOKLERRPVNKC	QDKLKN	TDPKKNNE	RESKKG	1260									
QY	1261	QRTSTFOIN	KDNKPKIYLGECLKEISESRVVS	GNPEPVNNINKI	IPENDIK	SLTVKE	1320										
Db	1261	QRTSTFOIN	KDNKPKIYLGECLKEISESRVVS	GNPEPVNNINKI	IPENDIK	SLTVKE	1320										
QY	1321	SAIRPFINGD	VIMEDFERNSSSETRSHLSS	SDAEGNYRDL	ETLPSTKES	SDSTQTTPS	1380										
Db	1321	SAIRPFINGD	VIMEDFERNSSSETRSHLSS	SDAEGNYRDL	ETLPSTKES	SDSTQTTPS	1380										
QY	1381	ASCPESNVQ	VEDMEIETSEVKKVTSSPTTSE	BSMLSNDP	IDENGLP	INKNENNGES	1440										
Db	1381	ASCPESNVQ	VEDMEIETSEVKKVTSSPTTSE	BSMLSNDP	IDENGLP	INKNENNGES	1440										
QY	1441	KRTVITEVT	MTSTVATESKTYIKVEGDKQYV	SVSTENCAKSTV	TTTTYTKL	STPS	1500										
Db	1441	KRTVITEVT	MTSTVATESKTYIKVEGDKQYV	SVSTENCAKSTV	TTTTYTKL	STPS	1500										
QY	1501	TGGSVDII	SVKEOSKTYVTVTVDLSLT	TGTLVTSMTV	SKEYSTRDK	VKLMFS	RPKKT	1560									
Db	1501	TGGSVDII	SVKEOSKTYVTVTVDLSLT	TGTLVTSMTV	SKEYSTRDK	VKLMFS	RPKKT	1560									
QY	1561	RSGTALP	SVYKFKYTKSKKSI	FVLPNDLKL	AKGIRFV	PFYNNAK	ALDIW	PPSP	1620								
Db	1561	RSGTALP	SVYKFKYTKSKKSI	FVLPNDLKL	AKGIRFV	PFYNNAK	ALDIW	PPSP	1620								
QY	1621	RPTFGIT	WRRLQTVKSLAGVSLMLRL	MAASLRMDMAA	KVP	PGGSTR	ETSET	ETITT	1680								
Db	1621	RPTFGIT	WRRLQTVKSLAGVSLMLRL	MAASLRMDMAA	KVP	PGGSTR	ETSET	ETITT	1680								
QY	1681	EIIKRDV	GYGIRFEYCIRKIIICPI	GVPEPTPKET	PPQKGR	SSALRP	KRETP	KQTG	1740								
Db	1681	EIIKRDV	GYGIRFEYCIRKIIICPI	GVPEPTPKET	PPQKGR	SSALRP	KRETP	KQTG	1740								
QY	1741	PVLIETW	AAEELIEMIRAF	AEVEKEKAQV	EQQAKRL	EQQKPLV	ITSTT	SPSST	1800								
Db	1741	PVLIETW	AAEELIEMIRAF	AEVEKEKAQV	EQQAKRL	EQQKPLV	ITSTT	SPSST	1800								
QY	1801	TSTISPA	QKVMVAPISG	SVTTGTRMVLJTV	GS	PAVTTF	FOONKN	FIQTF	ATWVKOGQNS	1860							
Db	1801	TSTISPA	QKVMVAPISG	SVTTGTRMVLJTV	GS	PAVTTF	FOONKN	FIQTF	ATWVKOGQNS	1860							
QY	1861	GVYVOQ	UGLII	PPSTGTSQ	QFTSFOP	RTAVTTR	PNPTSS	GSGTTS	NSQVIT	GPQIRP	1920						
Db	1861	GVYVOQ	UGLII	PPSTGTSQ	QFTSFOP	RTAVTTR	PNPTSS	GSGTTS	NSQVIT	GPQIRP	1920						
QY	1921	GMTVIR	PLQOSTL	GKAILTP	PMVOGAP	QVMTQIL	IRQAP	STAVAS	ANIV	SSIPGQ	1980						
Db	1921	GMTVIR	PLQOSTL	GKAILTP	PMVOGAP	QVMTQIL	IRQAP	STAVAS	ANIV	SSIPGQ	1980						
QY	1981	SLT	SATSTNS	TIQSSASQ	PPRPOGOV	KLTM	AOULTOL	TQHG	SGNOGL	TVLQOG	QOTTGOL	2040					
Db	1981	SLT	SATSTNS	TIQSSASQ	PPRPOGOV	KLTM	AOULTOL	TQHG	SGNOGL	TVLQOG	QOTTGOL	2040					
QY	2041	QILPQ	GVTVLP	PGGOOL	QMAAMP	NGTVOR	PLTAT	TAAT	TTTTV	STTAAG	TEGEOR	2100					
Db	2041	QILPQ	GVTVLP	PGGOOL	QMAAMP	NGTVOR	PLTAT	TAAT	TTTTV	STTAAG	TEGEOR	2100					
QY	2101	QSKLS	POMQV	HO	DKTL	PR	AOSS	SV	GR	AKAOPQ	TAOS	SARPO	QOTO	QOSPA	QOVQ	POVY	2160
Db	2101	QSKLS	POMQV	HO	DKTL	PR	AOSS	SV	GR	AKAOPQ	TAOS	SARPO	QOTO	QOSPA	QOVQ	POVY	2160

QY	2161	QOTTTVSSHVPSAEQPTPHAAOSSKPYVAAGOSPOPSNVOGOSPVVRVOSPSOTRIRPSTPSQL	2220
Db	2161	QOTOTTVSSHVPSAEQPTPHAAOSSKPYVAAGOSPOPSNVOGOSPVVRVOSPSOTRIRPSTPSQL	2220
QY	2221	SPGQOSQOVQTTTSPPIPIQHTLSLOIPSGQGQPSQCPQVQSSVQTLSSQGLNMOVSVSSPS	2280
Db	2221	SPGQOSQOVQTTTSPPIPIQHTLSLOIPSGQGQPSQCPQVQSSVQTLSSQGLNMOVSVSSPS	2280
QY	2281	RPQLOIQOPQPOVYAVNPQLOQOQVQVLSLOISQGVVAAIQAAQOSGVPQOIKLPLDIOQSS	2340
Db	2281	RPQLOIQOPQPOVYAVNPQLOQOQVQVLSLOISQGVVAAIQAAQOSGVPQOIKLPLDIOQSS	2340
QY	2341	AVQTHQIQIENVYVVOAASVQEQLOQRVOQLRBDQOQKKKQOQIEIKREHTLQASNOSEIIQO	2400
Db	2341	AVQTHQIQIENVYVVOAASVQEQLOQRVOQLRBDQOQKKKQOQIEIKREHTLQASNOSEIIQO	2400
QY	2401	VVMHNNAVTEHLKQKSKMTAEEREENOMLYCQVKKYTLDDIKDEEQAAKKKREESY	2460
Db	2401	VVMHNNAVTEHLKQKSKMTAEEREENOMLYCQVKKYTLDDIKDEEQAAKKKREESY	2460
QY	2461	EQRKSNQATKLSALTEFKHKEQRAEILKKRALLDDKDLOTEVQOEELKRDILKIKKEKDLQ	2520
Db	2461	EQRKSNQATKLSALTEFKHKEQRAEILKKRALLDDKDLOTEVQOEELKRDILKIKKEKDLQ	2520
QY	2521	LAQATVAAPCPVTPVLPAPRAPPPSPPPPCGVQHTGLSTPTLPVASQKKRREEKDS	2580
Db	2521	LAQATVAAPCPVTPVLPAPRAPPPSPPPPCGVQHTGLSTPTLPVASQKKRREEKDS	2580
QY	2581	SSSKKKKKMMSTTSKFKKQTKLYCICKTPYDESKRYICGDCRCQWMYHRCVGLIOSEAE	2640
Db	2581	SSSKKKKKMMSTTSKFKKQTKLYCICKTPYDESKRYICGDCRCQWMYHRCVGLIOSEAE	2640
QY	2641	LIDEVCPQOCSTEDATVTLPTLEKDYEGLKRVLSLOAHKMMAPTELEPPVPNDAPRY	2700
Db	2641	LIDEVCPQOCSTEDATVTLPTLEKDYEGLKRVLSLOAHKMMAPTELEPPVPNDAPRY	2700
QY	2701	GVIKPEPMDLATMEERVOARYEYKLETFEVAQMTKIFDNCRYNPSDSPFYQCAEVLESFV	2760
Db	2701	GVIKPEPMDLATMEERVOARYEYKLETFEVAQMTKIFDNCRYNPSDSPFYQCAEVLESFV	2760
QY	2761	QKLKGFKASRSHNNKLOSTAS 2781	
Db	2761	QKLKGFKASRSHNNKLOSTAS 2781	
RESULT 2			
Q95VB8 PRELIMINARY; PRT; 2669 AA.			
AC	Q95VB8		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Nucleosome remodeling factor large subunit NURF301.		
GN	EFBX) OR NURF301 OR CG7022 OR CG17135.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Empidoidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE-21468388; PubMed-11583616;		
RA	Xiao H., Sandelitzopoulos R., Wang H., Hamiche A., Rinaldo R., Lee K.,		
RA	Fu D., Wu C.;		
RT	"Dual functions of largest nurf subunit nurf301 in nucleosome sliding		
RT	and transcription factor interactions.";		
RL	Mol. Cell 8:531-543(2001).		
DR	EMBL: AF417921; AAL16644.1; -		
DR	FlyBase; FBgn0035118; E(box).		
DR	InterPro; IPR001487; Bromodomain.		
DR	InterPro; IPR000345; Cyt_c_heme_bind.		
DR	InterPro; IPR004022; DDT_dom.		
DR	InterPro; IPR000561; EGF-like.		

DB	InterPro: IPR001965; ZnF_PHD.	DR	Pfam: PF00439; bromodomain; 1.	DR	Pfam: PF02791; DDT; 1.	DR	Pfam: PF00628; PHD; 3.	DR	PROSITE: PS00633; bromodomain_1; UNKNOWN_1.	DR	PROSITE: PS00014; bromodomain_2; 1.	DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.	DR	PROSITE: PS01186; EGF_2; UNKNOWN_2.	DR	SEQUENCE 2669 AA; 300687 MW; 6B4925AFAF489D6F CRC64;	DR	SEQUENCE 2669 AA; 300687 MW; 6B4925AFAF489D6F CRC64;				
Query Match	19.8%;	Score	2639.5;	DB	5;	Length	2669;	Best Local Similarity	27.5%;	Pred.	No. 32e-105;	Matches	869;	Conservative	383;	Mismatches	919;	Indels	985;	Gaps	98;		
4	EEEEEDGAEETGTOSEDEDEMEEDDDSDSYPEEMEDDDASCTGSFSRSHSYSS	63	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152	99	ESEHYHGSFGSGSEEDKSINEDMDMLTPEDDESLEVANSSESEFSVC----	152
64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122	64	PEGRKKPRVHPRSPILIE-KDIPLEPKSSEDLVMPNHNINVAIYELRNFTVIR	122
153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209	153	---GRRPRPSPPEVWLGGRQYALDLPPSSSEDLFIANTHYLRALSIYEVLRRFRHMR	209
123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182	123	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	182
210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269	210	LSRPFREDCAALVDSOECTLAEMHNVLLKAVLRREDISNTTFGADLKDSVNSLYTI	269
183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238	183	DSMTPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	238
270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326	270	DSITMPEVPLRVVYCESDKEK-----HNHLPQZAEADYGVGEVKNKIKYLOFLVPOFLTNTA	326
239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298	239	REELMSEGVLOYDDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAAH	298
327	ROVMLOEGPIHVDHCRVCHLGLDCLCEFGSAVYHLCEVCKPRLEVPDEMOCEVCVAA																						

```

QY 889 VNCQESSQVDVNVYSEGPHL--RTSYKKTKTKSSKLDGLLBRRIKQPTLEEKORLEKIKLE 946
Db 837 -----IDVSKLNPARGRTIYSKAVAKRSRLDILLRKLAEVEE-QMAASKI--- 881
QY 947 GGIGKIGKISTNSSKMLSESPVITTKAKEGQSDSMROEQSPMANNDQPEDLIQCSQSDS 1006
Db 882 -----PSSDM 885
QY 1007 SVLRMSDPSTHTNKLYPKDVLDDVSTRSPETKCPKQNSIENDIEEKVSLASRGDEPTK 1066
Db 886 KPLIVSSQNTMANS-----KOTFILE-----KRLRLTE 913
QY 1067 SKTKGNDFEIDSKLASADDIGTLICKNKKPLIQEESDITVSSSKSALHSSVPKSTNDRD 1126
Db 914 VQAKG----- 918
QY 1127 APLLRAMEFGKLCQDSESNSTLENSSDTVS:IODSSEDMIVQNSNESISEQFRPREOD 1186
Db 919 -----GPAWVLELVNSLAKQIQTVRLQFS-----QL 945
QY 1187 VEVLPEIKCBLVGESEGNEDRLPVKGTANGKPKPSQCKLEBPVYNKCSDOIKLNTT 1246
Db 946 NREKAVFR-----YKKEC-----NTNSNVAVSQITQ-----NTCYSPLCLOKAR 984
QY 1247 DKRN-----NENRESEKKGQRTSTFOINGKDNKPKIYLKGECLKEISESRVSGNVEPKVN 1302
Db 985 AKKELLLLKKAHTAGNGSSETVAAILGAVKXPSIL-----EQKLTBEKRE--- 1030
QY 1303 NIKKIIPENDIKSLYKESAIRPFINGDVMEDPNEHNSSETSHLSSSDAEG----- 1356
Db 1031 -----STQVAVVDWSEEGKPAESAPDLIDLQDWEHARAHAV 1065
QY 1357 NYRDSL-----ETLPSTK-----ESDSTQTTTPSACSPESNSNOVEDM-----E 1396
Db 1066 PEDDSLITCELIWDQECVYTKIKQEVNASSGCGNTTPDSQTDSDKIDYIESMDVCSNVE 1125
QY 1397 IETSEVKVYTSPTSEESNL-----SNDFTID-----ENGLPINKNEN 1435
Db 1126 IESTESIVITGLNSGNAEDVDMTGPMRRKRNQSKSKSYIGTKDVLQDTLKDIPLNK--- 1182
QY 1436 VNESKRKYITVEYMTSTIVATESKTVIKVEKQKQTVASSSTENCACKSVITTTTTTVTK 1495
Db 1183 -----QNRREPIT-----ARPVKREC-----VKYERETFEENGNERV----- 1214
QY 1496 LSTPTGSGV-----DIISVKEQSKTVYTTVTDSLATTGTLVTSMTVSKEYSTRDKVKL 1551
Db 1215 YSTSSPRGRVYLLNDAKLYEQA-----VATEDKSTI 1246
QY 1552 MKTSRPEKTRSGTALPSYKREVTKSTKKSIFVLPNDLKKLAKKGITREVPYFNNAKPA 1611
Db 1247 TK--KESYSR-----YPLISFELTHKKRSLVLPREFELKLARLGSSSTNGFHAAKN 1300
QY 1612 LDIWPPSPPTFGITMRYVLYQVKSILAGVSLMLRLILMALSRLMDMAAKKPPGGSGTRTE 1671
Db 1301 -TITWQVQCSAPPLRTCKSYTNSATLSLALQDLRLIMLSGLMDMDMAK--PSTDGHQV 1358
QY 1672 TSETEITTEITLIRKRDVGPYGRFEYCIRKTIICPIGVPETPKETPTPOKRGKLSALRPK 1731
Db 1359 TITDTELITTELLKLRHSGRGKETSILRRKVVLPLEMPKRVREV--TSIRSGLR-----KRR 1413
QY 1732 RPTPKOTGVITETVAEELDELMEIRAFAEVEKEKAQAVEQO--AKKRLQOAKPTVIA 1790
Db 1414 RAESPQTEPQITEWVDEKDELMEIKFGEKQEKARLSAVTRVSAASROLE-----A 1466
QY 1791 TSTSTPSSSTSTISPAQKMNAP--ISGSYTTGTAKVLLTTKVSAPATVTFQOKNFHQTP 1849
Db 1467 SSGNSGNTSTNGALGVAGRQALPKLSEVYK--EKMEQOLK-----GRAVHQQR----- 1515
QY 1850 ATWVWQOGSNGVVOYQOAKVLG-----IIPSTGTS-----OOTFTSFO-----PRATVYTRP 1898
Db 1516 --LVAIGETIRSTYTPVKGQVIGSRVIVKKNPDGTTIRIQOAVTVQVSRGTGANTAAASP 1573

```

```

QY 1899 NTSQSGGTTNS-----QVITGPO-----IRPGMTVIRTP-----LQOSTLGKA 1937
Db 1574 TVGGSSTQSNPSTSTPHAKQIITIRGPDGKVSAGLNPQOOLVMPDGKHLVLTSTSSNS 1533
QY 1938 IIRTPVWVGAPQOVWTOIIRGQPVSTAVASADNTVSSPFGKSLTSANSTNSIQSSAQ 1997
Db 1634 AGQGNKKKVPKIPASTSS-----SPAISSAQTTNTPVPIKQIAVHVTKNSATQSIAS- 1688
QY 1998 PPRPQOGQVLTMAQLTQLTQGHGNGGLTVV:QGGQQTIGQ-----LQILPGGYV 2049
Db 1689 -----SRRVALPLAQI-----KNKLLLAQOQOQSTSSPANTSSPQKIVSKRVN 1733
QY 2050 LPPGQOLMAAMPNGVQVHFLEPTLATYATSTATTSTYTA--GTEGROQSKLSP 2106
Db 1734 TSTSGOTLQGVFVQSGS-----KLIVGQNNQGVKIIISTAAQOQGSIPVQOQOLVQ 1785
QY 2107 QMOVHODKTLPPAQSS--SVGPAKAQ-----PQTLQPSA 2138
Db 1786 SOPIQOS-----POQISMTQVGNQPTQKVIQIIVNTSNVQOQIIVGGORILISPGQITVWQ 1841
QY 2139 RPOQOTQPOSAPQEVQOTP-----EVQOTQTVSSHPSBAOPTHAKSSKPOVAAOS 2190
Db 1842 KNPQOSQALOMVQOQIQTQOQOQOHHVQPOQOPVVOSONIYVSSPSAQTKLYKQLVVOQ 1901
QY 2191 QPOSNVQOSPVVWQSPQST-----RIRPSTPQSLSPGQOQVQVQTTTQOP:PIQPHTSIQ 2245
Db 1902 QSOQTITEKTQIITTDSTENETGTOQVLPNSTLAQQLAGQLQVATVNGQOQVYKPLGNNQ 1961
QY 2246 -----IPSGOQ-----POSOPVQSSQTILSSQTLNQVSSPSRQLO 2285
Db 1962 AQIVAHKIKHGGDGNNAHVTSNSATVAPQANPOTSPVKQOALPPOSQOVVVO-----QOO 2016
QY 2286 IQOPQOP-----QVIAVPOLOQOVQV--LSQIQSO-----VVAQIOA 2319
Db 2017 IHQOSPINFESGVYPTIQOQVLTQVQVQOALQALSYEESLQONOPRGYIKVYTAQVQIO 2076
QY 2320 QQSG-----VPOQIKQLQLP-----IQIOQSSA 2341
Db 2077 TEHGPRIVLQGLVGNDFTAQQLQVQVQVQOQKMAQESNGKLGVLPTKIYLAQVQENA 2136
QY 2342 VQF-----HQIONVTVQA--ASVQEOQLQVQOLR-----DQOQKKKQOQ-- 2379
Db 2137 VQSQPPLPVPVHOSAAHQQANNIEIDADTLATYEVANSTKIDALINNGDOENSKAEVE 2196
QY 2380 -----IEIKREHTIQASNOSEILQOVYMKHNAVY 2409
Db 2197 NSNITTNESFAGTSSLSGSEHDEPTNLAGLOISEPDLQKQKESVYVVRGYIQK--AIS 2254
QY 2410 EHLKO-----KKSMTPAERENQMTV-- 2431
Db 2255 NALKQGNLSPELEKLVCMQKOQOENANSTNEMETCSRGSVNEBALTPRSQTDTEMKIRT 2314
QY 2432 -----CNOVMKYILDKIDKEEQAKKRRREESVQOKRSQKQATLSALFRK 2478
Db 2315 SLRPNAMTSSOPNKILK-----KNSKNDVAFELGQKQSO-----LER 2355
QY 2479 HKEQLRAEILKRRALLDKDLOIEVOELKRLDKIKKEKDLQALQATVAAAPCPVTPVYL 2538
Db 2356 HKELLKNNILKRRSLERLNQSEIHE--DVXTQVQHVRLPSNA-- 2397
QY 2539 PAPPAPPPPPPPVQVQHGLSTPVLV-----ASQKRRKEEK----- 2578
Db 2398 -----SPDEQSENERSG--EPNLDKFRTEVQNPBRHAGRPKLTTRKKEKLYXCIRT 2446
QY 2579 -----DSSKSKKKKMISTSKETKKDKIKVICICTPVD 2612
Db 2447 PYDDTKFYVQCDICSNWFHGDVCSITEEASK--KLSEFICIDCKRARETOQLCSCQOPVD 2505
QY 2613 ESKFYIGCDRCQNMVHGRGCILOSABEALIDEVVCQOCOSTEDBAMV--LTPILKEXYEL 2671
Db 2506 ESOFTYICDCKQDMFHGRGCILOSABEALIDEVVECCQRRKNDANNAKKLISNVEEL 2565
QY 2672 KVLRLSLQAHKMAFPLEVPDNDADPYGVIKREPMDLATMEERVQRRYEEKLTFEVADM 2731

```

```

Db 2566 KNLKIQMOLHRSAMPFEMPEVPKPEADYVKYIKPEMDLKRMEIKLESNTYTKLSEFIGDM 2625
Qy 2732 TKIEDNCRVYNPSDSPEFYQCAEVLSEFFVOKLKGFR 2767
Db 2626 TKIPDNCRVYNPKSESFYKCAEALESFYVQKIKNFR 2661

RESULT 3
Q8VDN7 PRELIMINARY; PRT; 645 AA.
ID O8VDN7;
AC O8VDN7;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hypothetical 72.2 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021489; AAH21489.1; -.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001965; ZnF PHD.
DR Pfam; PF00439; bromodomain.1.
DR Pfam; PF00628; PHD.1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO.1.
DR SMART; SM00249; PHD.1.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 645 AA; 72167 MW; 6486634F339820F9 CRC64;

Query Match 19.4%; Score 2775.5; DB 11; Length 645;
Best Local Similarity 84.6%; Pred. No. 2.2e-103;
Matches 559; Conservative 32; Mismatches 45; Indels 25; Gaps 10;

Qy 2125 GPAKQPOPAAPARQPOQPOSPAPQPEVOTQOTYTVSHVPEADQTHAOSKP 2184
Db 4 GPAKQPPAPQAQPPQ--QPQPPAQPEVQTP-----AVSHVPESTQPSQATSKP 55
Qy 2185 QVAASQPOSQVNGQSPVAVQSPQTRIRPSTPSQLSPGQSQVQVTTTSQPIPIQPHTSL 2244
Db 56 LVATQCQPOSQVNGQSPVAVQSPQTRIRPSTPSQVTPGQCPQVQVOTASQPIPIPPTSL 115
Qy 2245 QIPSGQPOSQPOSQVSSQVOTLSSQTLNQVSSPSRQLOLQOQPOVIANPQLOQVQ 2304
Db 116 QAPSGQPOSQPOSQVSSQVOTLSSQTLNQVYVLSQSP-----QPQPOVIAVQPL-QVQ 169
Qy 2305 VLSQSQSVNAQIQAGQSGVPOQIKLPQIQIOQSSAVQTHQIQNVYVQAASVQEOQLR 2364
Db 170 VLSQSQSVNAQIQAGQSGVPOQIKLPQIQVQNSAQT---QSVYVQAASVQEOQLR 226
Qy 2365 VQQLRDOQKKKQKQOQIEIKREHTLOASNQSEIIQKVYMKHNAVIEHLKOKKSMTPARE 2424
Db 227 VQQLRDOQKKK--QOIEFEREHTLOASNQSEIIQKVYMKHNAVIEHLKOKKSMTPARE 285
Qy 2425 ENQRIIVNQWYKTYLIDKIDKEKQAARKKRRESEVEQKRSQNAQTKLSALLFKKEQLR 2484
Db 286 ENQRIIVNQWYKTYLIDKIDKEKQAARKKRRESEVEQKRSQNAQTKLSALLFKKEQLR 345
Qy 2485 AEILKRALDLKDQIQVEQELKRDLTIKKEKLMQQLAQATAVAAPCPVPVLPAP-PA 2543
Db 346 AEILKRALDLKDQIQVEQELKRDLTIKKEKLMQQLAQATAVAAPCPVPVLPAP-PA 405
Qy 2544 PPSPDP--PPPGVQHTGLSTPT--LPVASQKRRKEEKSSSKKKKMTSTSKETKK 2599
```

```

Db 406 AAPAPRSPBPSTHSLPPAGPTAPLPTYSQKRKEEDD--SKSKKMTSTSKETKK 463
Qy 2600 DFKLCICIKTPYDESKFYIGCDRCQNMWYHGRVGLIOSEAEILIDYVPOCOSTEDAMTV 2659
Db 464 DFLKLCICIKTPYDESKFYIGCDRCQNMWYHGRVGLIOSEAEILIDYVPOCOSTEDAMTV 523
Qy 2660 LTPLTEKDYEGKLRVLRSLQAHKAMPLEPYDPNDADPDYGVIVKEPMDLATEERVOR 2719
Db 524 LTPLTEKDYEGKLRVLRSLQAHKAMPLEPYDPNDADPDYGVIVKEPMDLATEERIOR 583
Qy 2720 YYEKLTFFVADMTKLEPDCNRYNPSDSPEFYQCAEVLSEFFVOKLKGFRASRHNKLST 2779
Db 584 YYEKLTFFVADMTKLEPDCNRYNPSDSPEFYQCAEVLSEFFVOKLKGFRASRHNKLST 643
Qy 2780 A 2780
Db 644 A 644

RESULT 4
Q9W0T0 PRELIMINARY; PRT; 976 AA.
ID Q9W0T0;
AC Q9W0T0;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE CG17135 protein (Fragment).
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Holt R.A., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.D., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier F.,
RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Centler A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ijzerman C.,
RA Jastli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaRo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Part V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrnskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```



```

Db 1002 ----- 1001
QY 1428 LPINKENVNGESKRKVTITEVTMTSTATESKTVIKVEGDKQIVVSTENCASVTY 1487
Db 1002 -----SGVLGEKK----- 1009
QY 1488 TTTTYYTKLSTPSTGGSVDIISVKEOSKTVVTTVYDLSLTGTLTSMVSKVSTRD 1547
Db 1010 ----- 1009
QY 1548 KYKLMFSRPKKTRSGTALPSYRKFPYTKSKSIFVLPNDLKLKARKGIREV--PYEN 1605
Db 1010 -----AMPRIEIOFTSTKGGKSI FVLQKKILKQIMMGCGQGVYMGFS 1054
QY 1606 YNAKPAIDIMPSPRPFTFGITWRYLQTVKSLAGVSLMLRLMASLRMDMAKYPGG 1665
Db 1055 AGIKSNLLIMPYPAPRPTLDLCWKMQWLNARSLHAYALQKTIWSSIKRENF--PDDT 1111
QY 1666 GSTRTETSET--ETTTEITIKRDVGPYIGIRFECYIKRKICPI--GVPEYPERPTPQKRG 1722
Db 1112 HPDRRVVIDPSSHDERRRIRHKEMPEYGOYERYEMEIEIIPLYDEPEEDESMLSRNG 1171
QY 1723 -----LRSSALRPKRPE-----PKOT-----GPVIEFWVAEEBLE 1754
Db 1172 GSSEFSHRSSAKKRPQRHEFLSLKFGNYPKSKNAFSLDMNRKATATIRREWDGVYTK 1231
QY 1755 LWE-----TRAFAEREKEKAQAV-----EQQAKRLEPOOKPVIATSTTSPST 1800
Db 1232 VFEIKDWKWMIRAEAKTKRLEATRKAKADEBERRRIQQQQSVARIYV--PMHS-- 1289
QY 1801 TSTISPAKVMVAPISGSV-----TTGTKMVLTKVGSPTATVTOQ-----KNHQE 1849
Db 1290 ---LISEERNMV--PYLGSQOQRRPNNGENRGFLKYNSSSVSQAIGYASTPPGYHQP 1345
QY 1850 ATWVKQOSNGVYQOQKVLGIIPSTGTSQOTFTSFOPRTATVILRPMTSGSGGTSN 1909
Db 1346 PNIRQAGYNO-----LPKPTTSPNFOG--RP-VATITPTPOLRAAGADG-- 1390
QY 1910 SOVITGPQIRPG---MTVIRPLQOSTLGKAITRPVWQPGAPQVMTQILRGO---- 1961
Db 1391 --VVRVAMVMTPGKSTVNTNSTPYQ-----ALNRQOYQOQRQOQPAVARRLNGHFM 1443
QY 1962 -----PYSTAVSAPNTVYSTPGOKSLTSAT----- 1986
Db 1444 GTNRGGGRNPSYOMHRLDPNRAALORPGESETTEMRKRVTEALIPDNDGDEQPPVIRPY 1503
QY 1987 STSNIOSSASOPRPOGOVKLTMAOLTLOLGHGNGOGLTVYIGOGGOTGLOL---- 2042
Db 1504 PTSEFDAQRAQOQHPOSRPVYSTPAQMIKRTOPGVKH--NVILMASDGTOKMVLKPGQ 1561
QY 2043 IPQGVTVLPRGGQOLQOAMPRNGTVORLETPPLATTATTAATTTTSTVSTAAGTGEOROS 2102
Db 1562 FPPGVYI--STGQVVYVRQPTAVQOROLYT-----ATPGTFVVRILPANANGAPRQOH 1613
QY 2103 KL-----SPQMOYHOK--TLPPAQ-----SSSVGAKAKQPOTAQOSANP 2140
Db 1614 QVNRVVVQASGRPRAMEYMDQGTTPPGQOVRVYLGSGNSTGTPNVNPKVSSRGGPRGGLT 1673
QY 2141 QPOTOPSPAPQREVTQREYQV---QTVSSHVPSAQ 2175
Db 1674 MQMVQOQOQHNPRAHYMDPDATGFAVSTTTEQVPRDEQ 1711

```

```

RESULT 7
095208
ID 095208 PRELIMINARY: PRT: 1713 AA.
AC 095208;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE F26H11.2b protein.
GN F26H11.2b.
OS Caenorhabditis elegans.

```

```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT Investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 281515; CAC42289.1; -.
DR InterPro: IPR004022; DDT dom.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 1.
SQ SEQUENCE 1713 AA; 196824 MW; 178E470017D9AD5A CRC64;

```

Query Match 6.0%; Score 865.5; DB 5; Length 1713;
 Best Local Similarity 17.8%; Pred No. 1.2e-26;
 Matches 423; Conservative 298; Mismatches 613; Indels 1049; Gaps 73;

```

QY 4 EEEEBE-----DGAETODSEDE-----EDMEEDDDSDYPEMEDD---DDDASYTE 52
Db 169 KRQEDDIYMDDESEDEESSDEDEFMLNEDQVQVEEELNLTDIKIEGLDEENKCYC-- 226
QY 53 SFSRSHSTYSITPGRRKPRVHRPRPILBE--KDIPLEPPSSSESLAMPNEHINAVI 110
Db 227 -----PWLDEDPASLEKLEPSSODIPPTASIMDAVEI 261
QY 111 YEVLNKGTVLRSLSPREFDFCALYSQEQCTMAAHVLLKAVYREDTSTTPGPAD 170
Db 262 YELRSHYHTRILRTPTFEFDFCALISHNNSCTMAEYHVALRLNCLSDDEQTHSYTE 321
QY 171 LKDSVNTLYFIDGMTWPEVLRYCESDEYHNVLPYQEA-----EDYRGPYENKI 222
Db 322 TNSVNMIMHHMDLTLYAEILRQYLEA-----YFNADASVADAINVNYFVGVDANI 374
QY 223 K-----VLQFLVDQFLTTNIAREELMSGVQYQDDHCAYCHK--IGDLIC 265
Db 375 QRDSEFPFNKHYFRLVLLFLMSYRFLYSSEFKLVNNGKQONENCKVCSGSRVVG 434
QY 266 CETCSAVYHLECYKPRPLEVEPEDEMOCQEVQVAHK--VPQYDCAVEIQKKPPIRHEPIG 324
Db 435 CTQCEAAFHVEC--SHLKEPPE--VLVGNICCKNSAVRGVLPDEAVDREP--LRSQPIQR 489
QY 325 DRSRRKYWFLNRLTIEEDTENENEKKIYVSTKVQALAILDLCKDYWEAECLKILEM 384
Db 490 DRYGRYWFIVRLVQVSDETE---LYYSVLPQYLOLKLDRPTYEKDLCPTIRLR 545
QY 385 REEIHNMIDTDLTKKAGSNKSFLLAANEELIESIRAKK---GDIDNVKSP--ETEND 440
Db 546 IDEFLEQMALIVEMTSERR-----EALLETMYVRQDILGYDEAETATPQYLIHMD 594
QY 441 K--NETENDSKDAEKNREFEEDQSLKDSDDKTP-----DDPEQGSSEVGFSEKSSNG 493
Db 595 SKKRMASILRDAQOGVQVQVEKLEPRYVGQSPVACQVQVESILPESIGIF----- 647
QY 494 ELSSEPGAKGASGSTRITRILRNPDSKLSQOQVAAAHAHBANKLFKEGKEVLVNSQ 553
Db 648 -----DAKL-----IMFMSG----- 659
QY 554 GEISRLSTKKEYI-----MKGNIN-----NYFLKGQEGK---YRVYHNOYSTNSPLNKHQ 601
Db 660 -----ATQEELEVOFVIDSDNFDAPSANLIMRGEGNDQTFMYIYNYYSRNEMSESEFLT 713
QY 602 HREDHDKRHLNKHCLTPAGFEKW-----NGSVHGSKYLTITLTLLITTOLENNIPSSLT 657
Db 714 RKKADKKKKYMAKSF--AOIDNFVNVAKNROFYGDASLHCKFIWMTLOQVTKNPIIDLM 771

```



```

Db      6 NVEIESTEDSVTLGNSGNAEDVDMTPGWRKRNOKSKSYIGTKDVLDDGLDKDIPLNK 65
QY      1433 NEWNNGSKRRKTVITEVTMTSTVATESKTVIVIEKGDQOTVYSSSTENCAKSTVTTTTT 1492
Db      66 -----QNRFPIT-----ARPKREC-----VKYERETENGNERV----- 97
QY      1493 VTKLSTPTSGSV-----DIISVEKSKTVVTTVTDSLTGTGTLVTSMTSKEYSTRDK 1548
Db      98 ----YSTSSPRGRVYLLNDAKLEYEA-----VKTEDK 126
QY      1549 VKLMKSRPKKTSCTALBPYRKRFVTKSKTSFVLPNDLKLARKGCIREFPYRNVA 1608
Db      127 STITK--KPEYSR-----YPLISNFLTTHKKRSLVLPREFELLARIGSGSSTNGHHAA 180
QY      1609 KPRLDIPYSPRPRTGTITRYRYLQVTKSLAGVSLMLRLMASLRMDMDMAKAPPGGSGT 1668
Db      181 KNN-TTWOYCCSRPLRTCTSYTSNATSLSSIALDLRLMSCLRMDMDMAK--PPSIDGK 238
QY      1669 RTESETEITTEILIKRNDGVPYGIREFEYCIKRIICPIGVETPKETPPORCKRSSAL 1728
Db      239 HOYTDEIYVTLLEKLKRNHSGRYGKESYLRKRVVIPLEMPKTVREV-TSIRSGLR---- 293
QY      1729 RPKRPETPKOTGVYIETWVAEELEMETRAAEKREKKAQAVEQ--AKKLEDOOKPT 1787
Db      294 KRRRAESPOPEQITEEWDEDEKLEMETKEKGEKOKARLSAVYRSVASROLE----- 348
QY      1788 VITSTSTSPSTSTSTSPAKVWVAP-IGSVTGTGKAVLTITKVGSPATVTFPOOKNFH 1846
Db      349 --ASGSGNSTINGALGVAGVQIAPKLESDVK--EKMDOLKL---QAAVHQOR--- 398
QY      1847 QTFATWVKOGSNGVVOYQKVLG-----IIPSTGTS-----QOFTSFQ---PRATVPT 1895
Db      399 ----LVATGETIRSVTPVKGVIYSRRVIVKMPDGTTRIIQAAVQVSTGANTMAAA 453
QY      1896 IREPTSSSGTSS-----QVITGPQ-----IRGMVIRP-----LQOSTL 1934
Db      454 ASPVSGSTGTSNPSTSTPHKVOIIRPGKGVSVKRLNPGQOLVQMPDKLHVLTPTTS 513
QY      1935 GKAIIRPVWVQGAPOQVQTQIIRGQVSTAVSAPRTVSTFGOKSLTATSTSIQSS 1994
Db      514 SNSAGQGNKKKVPRIKPASTSS-----SPAISAQTTNPVPIVKQIARVHTNKSAAQSI 569
QY      1995 ASQPRPQOGVKILMAQTLQTLQGHGNOGLTVVIQGGQOTTGQ-----LQLIPOG 2046
Db      570 AS-----SSRVALPLAQI-----KNKLLIAQOQOOSTSSSPATSSSPVQKIVSK 613
QY      2047 VTIVLPGQOOLQMAAMPNGVQRFELTPLATATTASTTTVSTTA---GTGEQOSK 2103
Db      614 VVNTSTSGQTLQOEVQSGS-----KLIVGONAGQGVIIISTSAAGQGTSPVQOQO 665
QY      2104 LSPQMVQNHOKTLPPAOSSS-----VGPAAK--AOPQTAQPSARPOQOTQPOSAPQEPVQ 2155
Db      666 LVQSORIQGS-----PQGISMTQOQIIVVGGRIILSPQGITVTRQNVQSGALQVQOQIQ 721
QY      2156 TOP-----EVQOTQTVSSHVPESEAPRIHAOSSKQOVAAGOSQPOSQSGSPVRYQSP 2207
Db      722 TQOQOQOQHNHYVQPOQOQVVOQSNQIYVQSSPAQTKLVKQVLVVOQOQOQOQIETIEKQITITPDS 781
QY      2208 SQT-----RIKPSPTQOLSPGQOQSOVQTTQSOPRIQPHNSLQIPSGQGOQSOPOQVST 2262
Db      782 NETGTQOVLVNSTLQAOGLQAGKIQVATVNGQOVIYAP-----LNNQAOIYAHN 831
QY      2263 QTLSSGQTLNOVSVSPSRPOLQIQ--QPOPVIAPVQLOQOQOV--ISQIOSQVVAQIQAO 2320
Db      832 KHQDGNAMHVTSSATAVPAVQNPQSPVAKQALPQSPQOQVYVQOQOQHQQSPJNEESG 891
QY      2321 QSGVPOQIKL-----QLEPIQIOSSAVQTFHOIQN-----VTVQAASVQ----- 2359
Db      892 VTPITQOPVLTQAOVQAPAQ--QOALSYEESLQONQRPETIKVCTAYQVLQTEHGPRIVIOG 950
QY      2360 -----EOLQRYOQARDQOKKKQO 2378
Db      951 LVGNDFTAQOQLOQVOTQOVKQOOLMKAOE 977

```

```

RESULT 9
Q9W0T1 PRELIMINARY; PRT: 997 AA.
AC Q9W0T1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CG7022 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Galtbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Maasman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
DR EMBL: AE003467; AAF47362.1;
DR Flybase: FBgn0035118; E(bx).
SQ SEQUENCE 997 AA, 109285 MW, 8106060CFE6A8C7 CMC64;

```

Query Match 4.7%; Score 672.5; DB 5; Length 997;
Best local Similarity 26.1%; Pred. No. 3, le-19;
Matches 283; Conservative 132; Mismatches 343; Indels 327; Gaps 45;

```

QY      1394 DMEIEFSEKAKTVITEVTMTSTVATESKTVIVIEKGDQOTVYSSSTENCAKSTVTTTTT 1432
Db      6 NVEIESTEDSVTLGNSGNAEDVDMTPGWRKRNOKSKSYIGTKDVLDDGLDKDIPLNK 65
QY      1433 NEWNNGSKRRKTVITEVTMTSTVATESKTVIVIEKGDQOTVYSSSTENCAKSTVTTTTT 1492
Db      66 -----QNRFPIT-----ARPKREC-----VKYERETENGNERV----- 97

```



```

QY 195 CESDKEXHNVLPYOEADYRYPGVENKIKVLOFLVDQFLTNN----- 236
Db 164 NKGS-----CCPVRSYSDH-----ENELD-----FMDOSTTTTTTTPGFIILASTMP 209
QY 237 -----IARELMSEGV-IYDDHCYVCHKL-GDLIC---CE-----TCSAVYHL 275
Db 210 PTTTDCIHDDEIFADGSLGKNACHECHYCMRGDIYCAVOCECPMMAANGKSCRAMPAA 269
QY 276 ECVKPLPEVPDEMOCEVCAHVKVPTDVCY-----AELQNKPKYIRHPEIGYRKR 329
Db 270 EGECOSNYVCEDDSSTTEIYEITTPESATSVPAKGIHAIIPKEDVDLOHIDDNNKE 329
QY 330 KYWFLNRLIIEEDTENENKIMYSTKVOLAELIDCLDKDYW---EAEICKILEMRE 386
Db 330 TATIPSAELSGEIEEBEKO---KATVAPVQVTD--EKDPSFEPSSTAGIPSDSR- 383
QY 387 EIHRRHMDITDILNKARGSKSFLAANEILISIRAKK---GDDIDNVK- 432
Db 384 -----IDLPSSTSEESKESSTEAEDIVITVTPPEGSGEEDVPKPSQIPEKEIT 434
QY 433 -----SPEE-----TEKDKNET-----ENDSKD 450
Db 435 EDELIVKSTSAPAKASPEEVEVYATTSAPTEEDVKPTTACTISEEBEGKPTPAEBGSGE 494
QY 451 AEKN-----REFFEDQ-----SLEKSD-----DKTPDDDEOGKSEV 483
Db 495 EERDVKVTAPEETDEDAKPTSAVASDEKEQPKPSGSGDELDLKTATATAGATSA 554
QY 484 GPKSKSEKSNELSESP-----GAGKASGSTRITITLRNPDLSKLSOLKSOQA 531
Db 555 SESESEODECKSTEAPTSVDIEIPAKPTSESEBASGEDEVAKETTPAGASIASGEIEIV 614
QY 532 AAHEANKLFEKEEVLV-----VNSGEISRLST-----KKEVIMK 568
Db 615 KGTTPAGEPSEDEDEIVKTTPAEBSSESSEDELKVTTPPAGEPSVAGEELAKETTPA 674
QY 569 GININNTFKIGQEGKYRVYHNOYSTNSFALNKHQREDHDKRRHIAHKFCLTPAGEFKWNG 628
Db 675 GEPS-----IAGEEIVKV-----TPPAGESSIAGEEIVK-----VTPPAGESSSEG 717
QY 629 SVHGSVLT-----ISTLRILTQLENNIPSFILHNNASHR 665
Db 718 EEBITLVTTIPAGSSSEBGEIEYKESTPAGEPTESEBEDIKATTSAPASDIEGVKPEPT 777
QY 666 ANMIKA--VOMCSKPREFALALAILCAVAPVYMLPIWREFLGHTLRHMTSIEEREKEK 723
Db 778 ATEVPAEVEDFAKP-----TPIAEEBEPILAGTPIPIPDGIS-----GEEIEYK 822
QY 724 VKKEKKQEEBETMOATWYKTFPVYKHQVYKOKGEEYRVYTGWSWLSKTHVRFVPK 783
Db 823 GTTPQLEBOPELISEBTEV---PVAED-----D 848
QY 784 LPQNTVWV--YRKSLEGNKNMDEMDESDKRCSRSPKKIKIEPDEKND----- 831
Db 849 LSSSTSASAIASSTEGVODASTTTSAPARAGDKDEAATVYPTAODKODEEVEDATDL 908
QY 832 ---EVKGSAAKADONENDISKITEKKDQVAYELIDSDPKCKEPEMEVDMMK----- 884
Db 909 PVEDVYOSTTAKTTTEQKREESSTEARDAEIEVITTSPADQOEVEABADMDHDEED 968
QY 885 -----TESHVNCQESSQVDVNVVSGFHL--RTSYKKTKTKSSKLDEL 924
Db 969 VQATDPLPKSIDGIPVYVTEATTGQPETSDATDKPPSYLBPVSOGEVPSSTAKVDNR 1028
QY 925 LEERIKQFL-----EENQORLEKIKLEGGIKG-----IGKSTNSKSLSSPVIYTKA 972
Db 1029 NDEETKFTLPPSGEDOSSEPLPAMDLPAIGPEGDCLEVGKTYANNITVPATAPCDVSC 1088
QY 973 KEBCOSDKROGESPANNDOP-----DIJOGCSOS---DSSVLRMDSPTHTTKLVP 1023
Db 1089 K--CISISIVACQOMECKLEPNLEKCTVADDLIDGCCPTIICDESTSAEKDESTAK--P 1144
QY 1024 KDRVLDVSVISRPETKCPKONSTIENDIEK-----VSDLASRGQEPTRKSKTK 1070

```

```

Db 1145 DNKIDEDVSEISTE--ELPKVIMPTGTTEBPLSHVAPDERIQVTVTSVPAQFDSTTAKND 1203
QY 1071 GNDFFIDSKLASADDIGTLICKNKKPLIQESDPTVSSKSAHSSVPKSTNDRDATPL 1130
Db 1204 KKP--IDES-----ADKKRPIGESEED-----SKPIDESEEDKKPV 1237
QY 1131 SRAMDEBKIGCDSSESNTLENSDPTVSIODSSEEDMIV-----QNSGESISBQFRIR 1183
Db 1238 E-----ESAEKPKVDESEKPKPLVTVIPASEIEKESKPEDEKKT 1278
QY 1184 EOD-----VEVLELACELVSGESTGNCEDLPAVKGEANAK---KPSOOKLEERVWNC 1236
Db 1279 EADFAAPTEQPEATPAQIADTAKEVDOKLATNSAPVSEEDLKADEKKRE----- 1332
QY 1237 SDOI---KLKNTDKNKNEN---ESEKQORTS--FOJNGKDKPKIYKLEGLKEI 1287
Db 1333 TAOIPPAEIPASTDEPESSTELPTVDLKKPEEDSTKGEIAPESDKVPEVPTASSTENEI 1392
QY 1288 SESRVVSGNVEPKVNNINKLIIP--ENDIKSLTYKE-----SAIRPFIINGDVI----- 1332
Db 1393 EESDKFTTVAAPKISASDETEPTAEEDLVPAATFPIESEFEVSTKPAVOGPPPLPLARA 1452
QY 1333 -MEDFERNSSSEKSHILLSSDAEGNYRDLFTLPSTKSDSTOTTP--SASCPSNSVYN 1390
Db 1453 QPEKKPVDAETSTEADISTEPASAEVEKEASGETSESDNEIDAGASSTPVVASDEDEKTES 1512
QY 1391 QVEDMEIETSEYAKVTSPTT--SEESN---ISNFDIENGLPI-----KNENVNGE 1439
Db 1513 TEKTVE---ADKFTTVAPLAGDEESNLPKLIPODIFEEBA--PVAVTTAAPSDD---GE 1565
QY 1440 SK-----RKTVTEVTMTSTVAATESKTVYKVEKGDKOVVSTENCAKSTVT 1487
Db 1566 QKVEVEEKPIEDGQPIEDETSTPTS-----SENTEPESDRATTIAPSKEE--PSEFS 1618
QY 1488 TTTTIVTKLSTPSTGSSVDIIISYKESQKTVYTTVTVDLSLTGTLVYTSMTVSKYSTND 1547
Db 1619 TGAPTKDEPAEPSTDAPEDES--KETPESEVPTVAPA---GEKIPTSSITPDEBPTAT 1673
QY 1548 KYKLMKFSRPPKTRGSTALPSYKFTKSTKKSIFLPPNDLKLKLARKGIEVYFENN 1607
Db 1674 SAVAPKPDDEVEKETSTELPTDAPASSEDENS---STDQIPS-----EVP--E 1717
QY 1608 AKPALDIWYPSRPFTGILWRYRLQTVKSLAGVSLMLRLMASLRWMDMAKVPFG-- 1665
Db 1718 KRP-----ETPAQOTPEBGI 1732
QY 1666 -GSTRETSETITTEIIRKRDVGPYGIREFYCIARKIICPIGVPTPREPTPORK--GL 1723
Db 1733 VQATAAPPTSDEVPVQORLPE-----EVLAEIPOPSTETGI 1768
QY 1724 RSSALPKRPPEPKQCPVLIETWVAEELMEIAPAREYK--EKOAOVA-----Q 1776
Db 1769 KOODETTAAPSIDRK-----EPYVTEIDEATTVAPISKEKPKPEEKPEVQOKPTGEE 1822
QY 1777 AKRLEQOKPVYLAUSTSTSTSTSTISPAOKVMAVAPISGVTTGKAVLTTKVGSPTAT 1836
Db 1823 PSBEEKEKPIQDVBTEBPVSTSEASASTES-----SEEVKSTBEVAKEKED 1873
QY 1837 VTFQOKNKHQTPAIVWKGQOSNS--GVYOVQOKVLGIIPSSTG---TSQOTTSFOPTA 1892
Db 1874 -----KOPSSTAQAPVETIPEISITELPAQOGDKPTEBAPVDSDEDTSA 1916
QY 1893 TYTIR--PNTSSGGTISNSQVITGQIRKGMVYTRIPLOOSTLGAIIITPMAVPGAPQ 1951
Db 1917 PSDEKIPSVSG-----EEVEGPEV-----TTASPAOAEDEDLKTPAESSESSTD 1960
QY 1952 QV--MTOIIGQVPSTVAPSNVYSSPFGOKSLTSATST-----SNIOSSASQPPPOOQ 2005
Db 1961 KYPETTYOKPEBETRADELPEBESYTOVSVAITSTAPVAGGDIKEDQATTAAP--EEBEE 2018
QY 2006 VKLTMAQLTQLOGHGNQGLTVVIOGOGTTGOLQIPQGVTVLP--GPGQOLMQ--AAM 2062

```

Db 2019 IKPTIAPAAEIQ-----PSEKEPVDE-OEVESGKATPAEEDGCPIDEIAPA 2065
QY 2063 PNGTVQREFLEPLATTATTTTSTTAAGTGEBOROSKLSPOWQVODK-----TL 2116
Db 2066 TSGPIDE-----ASTAAPTKESTTVASA-----SP-AVHDEIKDVTITQ 2106
QY 2117 PPAOSSSGPAAQAPOTA-----OPSAR-POPOTOPOSPAPREVOTOREVOT--QTTV 2166
Db 2107 PVADKEVAAAPDETKTSIDVSTDSPTADDEKODTEAPVAPTTVSSPTAASADSSPT 2166
QY 2167 SSHVPS-----EAOPTH-----AOSKPOVA-AQOSPOSNVGGOSPVRR--OSPQTRIRPS- 2215
Db 2167 TVEVSPVEIDTRPMDLMSQTITAPTAGASTEDEDQAPVTVSPDAEKTVPSPAP 2226
QY 2236 -----TPSOLSPGQOSQOVYTTTQ-----PIPIOPHTSLQIPSGOGPOPOVOSTOT 2264
Db 2227 QOSDTPSEAPQADAEIPATATPIDDNKIPATVAPQTDGVPATAPRLDEDKIQTTAP 2286
QY 2265 L-----SSGQTLNOVSNSPSPRQLOIQPOPQVIAVPOLOOQOVLSQIOSQVAAQIQ 2318
Db 2287 LDEEKIPSTAAFLDEDEKIPAPVSPVDFVEPSESEKPAVSEYDGE-----ESTERPVDVE 2341
QY 2319 AAOAGVPOQIKLOLPIQIOSSAVQTHQIONVYVQA-----SVQEOLOQVQQLDDQO 2373
Db 2342 TSTDEPTSAKLKPPISAPATPSESAPATAEIVPETAPELEKEVEKATEQPELEKTP 2401
QY 2374 KKKOOIEIKREHTLOASNOSEI-----IQRVYMKHNAVIEHLKOKKSMTP 2420
Db 2402 EKATQPELEKETPEKATQPELEKETPEKATQPELEKEVDKATQPESEVDEKTTPEP 2461
QY 2421 AER-----EENORMIVCNOVMKYIILDKIDKEKQAKRRKRESEVQKRSQONATKLSA 2474
Db 2462 VVKPISLSTEEDE-----ESVSESEESADKKDKNKETEDTDKHEEPEVA 2508
QY 2475 L-----LFKHKEOLEKRAELIKKRALLDKDLQIEVQELKRLKIKKKE 2516
Db 2509 VVSEIPQSEEAVPPTGHPLEPH--LASTTTPPAVDR--VGEE-----DEBN 2553
QY 2517 DLMQLAQAATAVAPCPVPV-----PVLAPAPPPSPPPGVQHTGLSTPTLPVASOKR 2572
Db 2554 TVVKLSSTTTSTTESPVASSTTTVASOQOQPIPPRYG--HA-----PEDEY 2603
QY 2573 KREBKSSSKKKKKMISTSKETKKDKTKLYCICKTPYDESKFYIGCDRCOMWYHGRCV 2632
Db 2604 DEEEVFGPTCYAGKLYVSAQOIPRDPCDFCFC-----FRSDII 2644
QY 2633 GILOS-----EAELEIDEVCPQOQ-----STEDAMFVLTLP 2662
Db 2645 CLOQSCPPPIACGHEPISGFCPCRYECPVMAAVLINTTTSTTTLPLP 2695

RESULT 11
095YM2 PRELIMINARY: PRT: 17352 AA.
AC 095YM2:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE I-connectin.
CN I-CON.
OS Procambarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidae;
OC Astacidae; Cambaridae; Procambarus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
sarcomeres of crayfish claw muscle";
RL EMBO J. 20:4826-4835(2001).

DR EMBL: AB055861; BAB64297.1; -.
DR InterPro: IPR000577; FGGY_kin.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 48.
DR Pfam: PF00018; SH3; 1.
DR PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_3.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 4.0%; Score 572; DB 5; Length 17352;
Best Local Similarity 17.7%; Pred. No. 8.9e-14;
Matches 576; Conservative 570; Mismatches 1232; Indels 874; Gaps 133;

QY 3 SEEEEDDDDAETQSEDEDEDEDEDDDDSDSYPREMDEDDDDAATCT-----ESSFR 56
Db 12294 STEISDKKPKSEKATSIISQESISVQEISVKDAPQSIDAKPKTEQATSIISPHSLTI 12353
QY 57 SHSTVSTP---GRRKPRVHRPRSPILSE---KDIPLEPKSSEDLMPNEHIMNYI 108
Db 12354 QEVSVKETPTDISDRKPKTEKATSIISQESISVQEVSVKAGVETAKPKTEQATSI 12413
QY 109 AIVELRNFGVLRISPFREFDCAALVSQEOCT-LMAEMVVLKAVLREEDTSMTFG 167
Db 12414 SPHEISIQEISVKEAP---TDICDKPKSEKATSIISQESISVQEVSVKE-----A 12463
QY 168 PADLKSVMSTLYFDIGMTWPEVLKRYCESDKETYNHVLRYQAEEDRYGVENKIVLQF 227
Db 12464 PGSEIEAKPKT-----EYAKSSISPHSLTVQEV----- 12492
QY 228 LVQDFLTMLARELSEGVIOYDHCRCVCHKGLDLCETGSAVYHLQVCRPP--LEEV 285
Db 12493 -----LVKESAEISDKPKPOTGKATSIISQESISVQEVSVKAPGVEVY 12538
QY 286 -PEDEMOCEVCAVNAKVPVGYTDCV---AEIOKNKPYIRHEPIGYDRSRKRYFLNR-- 337
Db 12539 KPTQEQATSIISHQSLTVQEVTVKGDVSAEISDKP-----KTEKATSIISQES 12588
QY 338 LIIEPTENDENKIKIYVSTKQVLAELIDCLDKDYAEELCKILEMREPIRHMDITED 397
Db 12589 ISVQETVKEAPGSVVEDKLKPEQA-----TSVISPLESITVQEVSVKESSD 12636
QY 398 LTNKANGSNKSFLEAANFELLE---SIRAKGDIDNVKSPSE-----TEK 439
Db 12637 ISRRPKTEKATSIISQESISVQEISVKEAPGTADVAPKPEQATSIISHQSLTVQEV 12696
QY 440 DKNETENDSKDAEKRE---FEDQSLSE---KDSDDKTPDDP--EQGKS----- 481
Db 12697 TVKEIPTDISDKPKSEKATSIISQESISVHEVSVKDAFGSMKDAKPKTEQATSVISPH 12756
QY 482 -----EVGDF--KSEKSNGLS-----ESPAGKAGASTRI 511
Db 12757 ESTLVQEVSVREVPTEISDKPKSEKATSIISQESISVQEISVKAAPSGMSMDAKQTEQ 12816
QY 512 ITPRLRPDSKLSQKQVAAAHNAHNAKLFKEGKEVLYVNSQGEISRLTKSEVKIK--- 568
Db 12817 ATSVISPHSLTVQEVSVKEAPTELSDRPKPEKATSIISQESIS---VQEVSVKEAP 12872
QY 569 GNINNYFKLGQEGKYRVY--HNOYSTNSPALNKQHREDHDKRRLHAHFCFLTPAGEFW 626
Db 12873 GSVKD-LKLTQEQATSVISPHSLTVQEVSV--KEAPTEISDKK-----PKTE-- 12917
QY 627 NSGVSGSKVL---TISTRLRLTIT---QLENNIPESFLHPMKAS--HRAMWIKAYOMCSK 677
Db 12918 ---QATSIISQESISVQESVKGAPGSLKDEKPKSEQATSIISHQSLTVQEVTVKEA 12973
QY 678 PREPALALALBOCAVPPVMLPIWREFLGHTRLRHRTSISEREKE-----KYKK----- 726
Db 12974 PADIS-----DKP-----KTEKATSIISQESISVQEISVKEAPGSM 13011
QY 727 KKKQDEEE-----TMOQATWVRYTPPVKHQVWKQGEVRYVYGGSWISKT 775

```

Db 13012 KDAKPKTQATSVISPHESLTVQEVSVK---EVPEIISDKKPKSEKATSIILSEQESISV 13067
QY 776 HYRFRVPLPGNTNVNFKRSLEGTK-----NMMDNMDESKRKRSRSPKIKIETPD 827
Db 13068 HEVSVKSPSGMKDAKKTQATSVISPHESLTVQEVSVKSEPEIISDKKPKSEKATSI 13126
QY 828 -SEKDEVGKSD-AAKGADONEMDISKITEKKODVKEILLD--SDSDKPKCEPEMEVDMD 883
Db 13127 LSEQESISIQEVSVKEVPEESMKDLKPKTEQATSVISELISLTVQEVSVKSEPGLSDKK 13186
QY 884 KTE---SHVNCQESSQVNVVNSGPFHRTSYKKTK--SKLDGLERRIKOFTLEKQ 938
Db 13187 KTEKATSIILSEQESISVOEISVKEAPGSVEEVKPKTEQATSIISPHMSLTVQEVTKE-- 13244
QY 939 RLEKIKLEGIGIKIGIKTSTNSKNLSSEPVYTKAKEGQOSMSMQEOQSPANNDQPDLI 998
Db 13245 -----VPADISDVKPKSEKATSIILSE-----QESISVOEISVKEAPGSMD----- 13285
QY 999 QGCSOSDSVLRMSDPSSHHTNKLYPKDRLVDVSIRESPEKCPKQNSIENDIEKVSDLA 1058
Db 13286 -----AKPKTEQATSVISPHESLTVQEVSVKSEPEIISDKK 13321
QY 1059 SRGQEPKSKTKGNDFFLDOSKLASADDIGLICKNNKPLIOESDPTIVSSKS-----A 1113
Db 13322 PKSEKATSIILSEQESISVOEISVKEAPE---SKMDAKPK-TQEQATSVISPHESLTVQEV 13376
QY 1114 LHSSVPKSTNDRDATPLSRAMDFEKLGCDSSENSLTENSDDVYSIQDSEEDMIQVNSN 1173
Db 13377 SVKEVPEIISDKK-----KTEKATSIILSEQESISVOEVS---VKAP 13416
QY 1174 ESISEQRTHEQVENVLEPLKCELVSESGNCEDLPLVKGTEANGKPKSOQKL--E 1229
Db 13417 GSVADLKPKEQATSVISP-----HESLTVQEVSVKEAPTEIISDKKPKTEQATSVILE 13469
QY 1230 ERPVNCSDQIKLNTDKNNENRESEK-----KGQRTSTPOJNGK-----DNKPK 1276
Db 13470 QESIS--IOEISVKEAPGSLKDEKPKSEQATSIILSEQSLTVQEVYKAPADISDKPK 13527
QY 1277 IYLGEBLKE-----ISSRVSGNVEPKVNNIKIIPENDIKSLTVKESAIRP 13326
Db 13528 TEKATSIILSEQESISVOEISVKEAPGSMKDAKPKTEQATSVISPH--SLTVQEVSVR-- 13583
QY 1327 INGDVIMEDENRNSFEKSHLSSDAEGRNDLSLETLPSTIESSTQI--TTPASACP 1384
Db 13584 ---EVPEIISDKKPKSEKATSIILSEQESISVOEISVKEAPGSMKDAKPKTEQATSVISPH 13640
QY 1385 ESSNSVQVQEDMEIETS---EVKKVTSPTTSEESNLSDNDFIDENGLPINKNENNGES 1440
Db 13641 ESILTVQEVSVKEAPTELSDEPREKEKATSIILSEQES--ISVQEVSVKEAPESKAPTE 13699
QY 1441 KRTVITEVYTMSTVATESEKTYIAKEKGDQOTVASTENCASKSTVYTTTTVTKLSTPS 1500
Db 13700 QATSVISELISL-----SVQEVSVKSEPEIISDKKPKTEKATSIILSEQESISVOEISVKE 13754
QY 1501 TGGSVIISVYK-BOSKTVVYTVTDLSLTTGTL-----VISM-----T 1538
Db 13755 APESSVEVVKPKTEQATSIISPH--SLTVQEVTVKGPVADISDVKPKSEKATSIILSEQES 13812
QY 1539 VS-KEYSTQDKVYLMKFSRPKTRSGTALPYRKFTVTKSTKKSIFLPPDDLKLARKCG 1597
Db 13813 ISVQEVSVKEAPESKMDAKPK-TQEQATSV-----ISPHESL--TVQEVSV 13853
QY 1598 IREVPFNNNAKPAIDINWYPSRPFGITWYRIQTVYKSLAGVLMRLMLWASLRWMD 1657
Db 13854 VREVPPEIISDKKPK-----KTEKATSIILSEQ-----ESISVOEI 13886
QY 1658 AAKVPPG--GSTRTESETEITTTTEIKRHDVGPIGRFPCYLRK----- 1701
Db 13887 SVKEAPGSMKDAKKTQEQATSVISPHESLTVQEVSVKEVPEIISDKKPKSEKATSIILSEQ 13946
QY 1702 ---IICPIGVPETPK-----ETPTPQKGLRS--SALPKRPETPK 1737

```

```

Db 13947 ESISVOEISVKEAPESKMDAKPKTEQATSVISPHESLTVQEVSVKELPEIISDKKPKTE 14006
QY 1738 QGCVIIEETVWAELELMEIRFAERVEKEKQAAVEQAKRLEDOQKPTVIASTTSP- 1796
Db 14007 ATS-ILSE-----QESISVOEISV-----KEAPGSMKDAKPKTEQ-----ATSVISPH 14048
QY 1797 -----TSSITTSISPAQKVMVAPISGVTGTGMVLTKVGS 1833
Db 14049 ESILTVQEVSVKEVPTDISETKPKTEKATSIILSEQESISIOEVSVKBATG-----SKMDK 14103
QY 1834 PATVPEONKNFQTEFATWVKQGSNGVVOVQOKYLGIIPSTSGTOFTSFORFAT 1893
Db 14104 P-----KTEQATSVISELIS--LTVQEVSVKSEPEIISDKKPKTE 14142
QY 1894 VTIRPNTSGSGGTTNSQVITGQPIRGMTVIRTPLOOSTLGKALI--RTPVWVQGAPO 1951
Db 14143 AT-----SILSEQESISVOEISVKEAPCSMEVVKPKTEQAT--TIIOSHESLTVQEVAVK 14195
QY 1952 QVMTQIIRGQP---VSNVAPNTVSTPGOKSLTATSNTSQSSASQPPRPOQOVKL 2008
Db 14196 EVPTDISDKKPKTEQATSVFSP-----QESISVOEVSVKEAPGAVKDSKPKTEHAKS 14247
QY 2009 TMAQLTQLFGHGNQGLTV-----VIOGQGGTTGLOLIRQGVTVLRPGQOLMQAAM 2062
Db 14248 TISPLESLTV-----QETVYKESPRDIEELKPKTEKATSVILEQESI--SVEEILVKDV 14299
QY 2063 PNCTVQRFLETPLATATATSTTTTVSTTAAGTGQROSKLSPOMQVHODKTLR---P 2118
Db 14300 PGSMDLKPKTEQATSVISPHOSLTVKEITVKEETPEIDLDE-KPRTIELAKSILLPODSIA 14358
QY 2119 AOSSVGPAKCAQPOTAQAP---SARPOQOTQOSPAPQEPOTQEPVOTOT----- 2165
Db 14359 VQEVSVKEAPVDIKELKPKSEKATKAPVYSTQ--ESLSVEVQVGOQVQEWMTPKQISETARV 14417
QY 2166 -----VSSHV-PSEAPQTHAOSKPOVAAQSPQSNVQGSFVRQSPS--- 2208
Db 14418 SYTKKDSQVLEVSSSELADIDIPETRLIEDKAKVYTEMILQABDQ--EDVIYVLRGPTKEK 14476
QY 2209 -----QTRIRPSTPSQ-----SPGQOSQV-Q 2229
Db 14477 IIEVSESEBELVKEIPEYKEEIEVEQPKLKPKTKKKEITTEVYIIRKEEVEIE 14536
QY 2230 TTTSGPIPIQPHTSLOIPSGQOPQOPQVOSTQ--TLSSGGTLQVNSVSSPSR----- 2281
Db 14537 EVIKRPDEIEVVEVWIKKPKKTEVTTEVEVVEVYIIRKEEVEVYVKKPEKIEBIT 14596
QY 2282 PQOLOQQPQVYIAPPOLQOQVQVLSQIOSQVVAQIOAQOSVPOQIQLDLPQIOQSSA 2341
Db 14597 EEVYIIRKPEEKV--VEEITTEVQIIRKPKKPKPEKIEE-----VVEEVLTKRP-----EEK 14645
QY 2342 VQTHOIQWNVTVQAASVOEOLQORVOQLRDOOQKKQOOQI-EIKRHTLQOASNOSEIIQO 2400
Db 14646 IAEVYTEELVYVAKAEKVEVYEEVQIIRKPKKPKPEIIEVYTEVYIIRK--KPEEYTEE 14702
QY 2401 VVMK-----HNAVIE-----HLKQKSMTPAER-EENORMIYONQ 2434
Db 14703 VYIKRPEEVYHEVSEVYIIRKPKKEVVEVYEEVQIIRKPKKPKPEIIEVYTEVYIIRK 14762
QY 2435 VVKYITLIDIDEKQOAAKKRR-----DESVQKRSKONATKL-----SALLKHKHEQL 2483
Db 14763 PEEKVEEVEVTEVYIIRKPKKPKPEKIEEVEVYIIRKPKKEVYEEVYEEVYIIRKPEEKV 14822
QY 2484 RAEILKRRALLDKDLOI--EVOEELKRDLLTKK-----EKDMLQOLA 2522
Db 14823 HEVSEVYIIRKPKKEVYEEVYIIRKPKKPKPEKIEEVEVYIIRKPKKEVYEEVY 14882
QY 2523 QATAVAP-----CPVTVPLPAPPA-----PPSPPPPGVOHTGLSTP 2563
Db 14883 EEVYIIRKPEEKVEVSEVYIIRKPKKEVYEEVYIIRKPKKPKKPEIIEV-----TE 14938
QY 2564 TLPPVASQKRRKEE-----KDSKSKSKKKKMTSTTSKET--K 2598
Db 14939 EVYIIRKPEEKIEEVEVQIIRKPKKPKPEKIEEVEVYIIRKPKKEVYEEVYIIRKPEEKV 14998

```



```

Db 3667 LVGQLSPVGRWVSP--SPQO---QVOQOTQOQNALITSPQSSNISPLASPTTFLVSSSNS
QY 1992 QSSA---SOPPRQGGGVKLTMAQLPQLTGHGNGOGLTVVIGQGGTQOLQILPCGV 2048
Db 3723 PTTSKVNSVGRNDQ-----VPOSPKSAVEVOTTPQLMT
QY 2049 VLPEPGQQLQMAAMPNGTVGRFLPTPLATTATATSTTTTSTTAAAGTGROSKLSPQM 2108
Db 3759 I---PLOKMPRIQVPHNPT---ITSKVTVVQPOQATQSVASSPRLGS-----LPRHK 3805
QY 2109 QVHODKTLPRPAQSSSVGRPAQAQPT--AQPSARQPTQOPQSPAPQREYQTOPEVQOTTVS 2167
Db 3806 NVH---LNAHQN-----QOQPOVIAKMTAHQHOQMFHQMOIMIOROHQMO----- 3849
QY 2168 SHVPEAQPTHAQSSKROVAQASQPSNVQSGSPVTVQSPQTRIRPSTPSQLSPQO-QS 2226
Db 3850 -----QOQLHGQSO--QITSAPOHOMHOQHQOQOQHNNQOHLMOQLHAQOHPQOKH 3901
QY 2227 QVOTTTSQPIPIQPIHTSLQIPSGQPOPOVQVS---STQTLSSQTLNVQSVSPSRPQ 2283
Db 3902 QAOQOQFQO--QIQOHOQSOQ---QHVOQOQOQAOQOHLQOQHQSGOQQLNQ---QHQAQO 3952
QY 2284 LQIQQOQOP--QVIAVQPOLQOO-----VQVLSQIQSOVVAQIQAOQSGVPPQITK 2329
Db 3953 QOQLOQIQKLOQMHGPOOQOQSPQGVGHGSGTSTIFASQOHNQOLPAR-----GVPOQ-- 4004
QY 2330 LQLPIQOQSSAVQTHQIQNV--VTVOAASVQEOQLQRY--QQLRDQOQKKQOQIEIKREH 2386
Db 4005 -QHPOQLSHSSPCKPMTIVSVNGVQPPAL---LTVGSHSQPNQOQOLPHOQSSGHPH 4060
QY 2387 TLQASQOSELQIKO---VVMKHNAVI--EHLKQKSKMTPAEREQNMIVCQNVKYL 2440
Db 4061 QKQLOSPGANLPQOTPLNYIQMTPKIIVQOHIVAQNVPPQPGQNAIHYPQNO--GKOST 4119
QY 2441 DKLDKEKQAAKKRKEESEVEQKSKONATKLSA---LFRKKEQLRAEILKKRL 2493
Db 4120 PRCHVPTPTAMSAQKTSSESVYKRTPTPTGLAVISANTVGSILTEENLIKISQPKODEL 4179
QY 2494 LDKDLOIEVO-----EELKRLDIKIKK-----EKDLMOAQATVAAPCPPTVLPAP 2541
Db 4180 IEQDSK-EVDSQVSAKEVINIVSIKIKDTPLASKDKAKRAVEQAL----- 4224
QY 2542 PAPPPPPPPGVQHTGLSTPLPVAQSKRRKEEKQSSSKSKKKM 2569
Db 4225 -APAPIPNPQPGNO--SMAQETALPTTSMNVNSNDHDEDETEFQOL 4269

RESULT 14
0906C3
AC 0906C3; PRELIMINARY; PRT; 5533 AA.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE SPEN RNP motif protein long isoform.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOL;
RA Wielelte E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
RA McGinnis W.;
RT "Spenn encodes an RNP motif protein that interacts with Hox pathways to
RT repress the development of head sclerites in the Drosophila trunk.";
RL EMBL; AF188205; AAF13218.1; -
DR HSSP; P09651; IHA1.
DR FlyBase; FBgn0016977; spen.

```

```

DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.
SQ SEQUENCE 5533 AA; 597114 MW; AFG606060DEF269 CRC64;

Query Match
Best local similarity 19.2%; Pred. No. 7.1e-14;
Matches 520; Conservative 379; Mismatches 931; Indels 878; Gaps 118;

QY 325 DRSRRKYWFLNRLIIEEDTENENKKIWIYSTQVLAELIDCLDKYVAELCKILEEM 384
Db 1997 DKQEKKEIREDKDLREKQRENDREK-----ELRDKDLREKEM-REKEQR 2041
QY 385 REEIHNMITE-----DLTKAGSKNSFLAANEELISIRAKKGDIDNV 431
Db 2042 EKLHREKQREHREKREKQSRRAMDVEQGGGRM-----RELSTYQKSKMDIAGE 2093
QY 432 KSPETEEDKNETENDSKAEKNREFEEDQSL--KSDDKTPDDPEQKSE----- 482
Db 2094 AS-----SLTADICQHNKENAMDTIAQGTGASPTSPDPTPKERSKLSRNSP 2142
QY 483 VQDFSEKSNGLSESPGAKGASGSTR-----ITRLRNPDSK-----LSQLSQOV 530
Db 2143 VRLHKRLSSQESNHSAGCGGSGSHQIHNEQYKRLIMENSONISVHNSQRLNDR 2202
QY 531 AAANHAANKLFKEKGEVLVNSQGEISR-----LSTKK----- 563
Db 2203 DSKENHSSS-FKEDK-----NSSHTSRPHGCGGSASSSKHHNRDKHHQGSASSIET 2256
QY 564 ---EYIM-----KGINNYFKLGQ-----EGKYRVYHQYSTNSFALNK-----HOHR 603
Db 2257 NSIEVVDPIQOTHNNTSEELQSHQPKREKREHNSHANSSSRHKSKRPHNHR 2316
QY 604 -----EDHDKRHLAHFCLTPAGEFKWNSVHSGSVLTISTLRLLTIOLE 649
Db 2317 EKRHSVASTNTDEHPQOHNPHR-RISAAG-----SGSAG-----ELSSAA 2359
QY 650 NNIPSEFLPHNASHRAWNIKAVQCSKPREFALALILECAVKNPVMLPIWREFLGTR 709
Db 2360 TTSTSGKLHHQ--HHRVSVERKSSRGSDGHHSSKSL-----RAK 2398
QY 710 IHRMTSIEREEK-----EKVKKRKKQEEETMOATVWKYTF 747
Db 2399 LMLSSASDDTDADSKKHSIFDIPDCPNVSMYKVABQSKNNQRAEKKIKAKRS- 2457
QY 748 PVKHQVWKQGEERYVYTGCGHS---WISKTHVYFVFKLPGNTNVNRYKSLGKTKNMD 804
Db 2458 ---QLKQSRRAKKRSTSYDGDSDTEFEDRQH-----RNSGSSSFHGR--YFGLSSSD 2505
QY 805 ENMDESDRKCGRSPKIKIIEPDSEKDEVKSDAKKADADQNE-----MDISKITEKKQD 859
Db 2506 DDEDETHORRIS-----SSDAHGQDQNGQSTLADANRVRQMO-QN 2547
QY 860 VKELIDSDSKPCKEPEVVDMDKMTESHVNCQESSQVYVNVSEG-----PHLRTSY 912
Db 2548 LRLICDGDSDS---SPD--EIRRNVMKHSHPGRKNSNSTRIASDSBSQSPAPADLTIOEH 2602
QY 913 -----KKTTSKSLDGLERKQTLLEKQRLKYLEGGIGKIGKSTMSKNLS 964
Db 2603 PIAPQEIIRQQLSDEQKFKSRHDSNSIEER---KLKTEREI---KTELQDFYNSS 2654
QY 965 ESPVITRAKEGQSDSMQEOQSPNANNQDPPDLIOGQSQSDS-----YLRMSDPs--- 1015.
Db 2655 EYITGKLKLEISPEPRKKHKKSKRR-----IKSSSTADTSAQAQPLVMTPLTPSTFD 2706
QY 1016 -HTTNKILVPKXRDVLDVSI-----RSPETCKPKNSIENDIEEKVSLDAS 1059
Db 2707 VHSSECKTFKDNPDLDKTECCSIPLEISAGERKKHKKRKKKREKLRNMTFAIVP--- 2763
QY 1060 RQGEPTKSKTGNDPFFIDSKLASADIGITLCNKKKPLIOESDITVSSSKALHSSVP 1119
Db 2764 ---SPTTNDT-----SSKLSKEERHRLKSKSKSKSMQNSCNTKIYNSSGA-H---P 2808

```



```

Db 3717 PTTSKVNSIQPRNOQ-----VPOQSPKSVAEVQTTPOLMT 3752
QY 2049 VLPBGQOLNQAAMPNGTVORFLTPLATTATVSTTTTSTTAAGTGEOROSKLSBOM 2108
Db 3753 I---PLOKMTPIQVPHNPT---ITSKVTVQROQATOSQVASSPRLS-----LPPHK 3799
QY 2109 QVHODKTLPPAOSSSVGRAPAKQPT-AQPSARPOPTOPQSPAPREVOQTOPEVOQTTVS 2167
Db 3800 NVH---LNAHON-----QOOPVIAKMTAHNOQHMOPMHOQMIOROHMO----- 3843
QY 2168 SHVSEAPQTHAOSKPOVAQAOSPOSNVQOSFVRVQOSPSQTRIRSTPSQLSPGQ-QS 2226
Db 3844 -----QOOLHGOSQ-QITSAPQOHMOQOQAQOQOQHNQOHLNOQLHAQHPQKH 3895
QY 2227 QVOPTTSQPIPIQPHNTSLQIPSGQPOSOPQVOS---STQTLSSGOTLNOVSVSSPSRPQ 2283
Db 3896 QAOQOQFWQ-QIOQHOSQ---QHQVOQOQNAQOQOHLSQOQHSQOQLNQ-----QHQAQO 3946
QY 2284 LQIOQPPQ-QVIAVPQLOQO-----VOVLSQIOSQVVAQIOAQOSGVPQOIK 2329
Db 3947 QOLQOIQKLOQMHGPOQOQKSPQGVHIGGSTSIFASQOHNSQLPAR-----GVPOQ-- 3998
QY 2330 LQLPQIQOSSAVQTHQIQNV-VTVQAASVQEOQLQRY--QOLRDQOQKKQOQIEIKREH 2386
Db 3999 -QHPQOLSHSSPCKPMTLVSNOGVOPPAI--LTRVGSHPQPNQOQOLPRHOSSSGPH 4054
QY 2387 TLOASNOSEIIQKO---VVMKHNAYI--EHLKQKSMTPAREENQRMIVCNOVMKYL 2440
Db 4055 QKOLSSPGANLPLQTPNLNVIQNTPKIIVQOHIVAQONOVPPQTOGNALIHYPQNO-GKDST 4113
QY 2441 DKIDKEKQAKKKRRESEVEOKRSKONATKLSAL-----LFKHKEQLAEILKKRAL 2493
Db 4114 PPGHVEPTPAMSAQKTSSESYSVIRPTPTGLAVISANTVGSLLTEENLIKISQPKDEL 4173
QY 2494 LDKDLQIEVO-----BELKBDLKIKK-----EKDLMQLAQATAVAAPCPPTVPLPAR 2541
Db 4174 IEQDSK-EVDSDYWSAKAEVINDSVIKKLDTPRLASKDAKRAVEMQAI----- 4218
QY 2542 PAPPPPPPPPGVQHTGLSTPLPVAQOKRKRREKSSSKKKKK 2589
Db 4219 -APAPTPNPQPGND--SMAQETALPTTSMGVNNSNDHDTEDETETROL 4263

```

Search completed: November 20, 2002, 16:33:47
 Job time : 315.851 secs

SQ SEQUENCE 810 AA: 91799 MW; F7E2C992FE5BE96D CRC64;
 Query Match 23.5%; Score 3363; DB 1; Length 810;
 Best Local Similarity 82.6%; Pred. No. 2.5e-113;
 Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2;

QY 1 MWSSEEEEDGAAETQSDSEDEDEDEDDDDSDYPREMEDDDDDASYCTESSFRSHST 60
 |||||||
 DB 1 MWSSEEEEDGAAETQSDSEDEDEDEDDDDSDYPREMEDDDDDASYCTESSFRSHST 60
 |||||||
 QY 61 YSSTGRRKPRVHRPSDILEEKDIPLEPKSSSDILAVPMHIMNVTAIEVLRNGETV 120
 |||||||
 DB 61 YSSTGRRKPRVHRPSDILEEKDIPLEPKSSSDILAVPMHIMNVTAIEVLRNGETV 120
 |||||||
 QY 121 LRLSPFRREDPCALVSOEDCTLAAEMHVLKAVLRREDTSNTTFGPADLKDSVNSTLY 180
 |||||||
 DB 121 LRLSPFRREDPCALVSOEDCTLAAEMHVLKAVLRREDTSNTTFGPADLKDSVNSTLY 180
 |||||||
 QY 181 FIDGWTPEVLRVYCESKKEVHVLPTQEAEDYRGPVENKIKVLOFLVDFLTNTIARE 240
 |||||||
 DB 181 FIDGWTPEVLRVYCESKKEVHVLPTQEAEDYRGPVENKIKVLOFLVDFLTNTIARE 240
 |||||||
 QY 241 ELMSEGVLYQDHCNVCHKLDLCCETCSAVYHLECVKPLEEVPEDEMOCCEYCAHKV 300
 |||||||
 DB 241 ELMSEGVLYQDHCNVCHKLDLCCETCSAVYHLECVKPLEEVPEDEMOCCEYCAHKV 300
 |||||||
 QY 301 PGVNDCAVEIQRNKPRTHEPTGDRSRKKYWPINRLLIEEDTENENEKIKWYSTKVQ 360
 |||||||
 DB 301 PGVNDCAVEIQRNKPRTHEPTGDRSRKKYWPINRLLIEEDTENENEKIKWYSTKVQ 360
 |||||||
 QY 361 LAELIDCDKDYMAELCKILEEMRETHRMIDTETJTKARGSNKSFLAANEETLES 420
 |||||||
 DB 361 LAELIDCDKDYMAELCKILEEMRETHRMIDTETJTKARGSNKSFLAANEETLES 420
 |||||||
 QY 421 IRAKKGIDNVKSPETEKKDNETENDSKDAEKREEFEDOSLEKSDDKTPDDPEQK 480
 |||||||
 DB 421 IRAKKGIDNVKSPETEKKDNETENDSKDAEKREEFEDOSLEKSDDKTPDDPEQK 480
 |||||||
 QY 481 SE----- 482
 ||
 DB 481 SEETPEVGDKNSVANLGDNTNATSEETSPSEGRSPVCLSETPDSNMAEKVASEL 540
 ||
 QY 483 ----- 482
 ||
 DB 541 PDVPEEPNKTCESSNTSATTTSIQPLVLENSNSSSELSNSESASAKAADDPENGERSHT 600
 ||
 QY 483 -----VDFKSEKNGELSESPPGAGKAGSGSTRITRLRNPDSKLSQKSOVAAAA 534
 |||||
 DB 601 PVSIOEBIVDFTSEKSTGELSESPGAGKAGSGSTRITRLRNPDSKLSQKSOVAAAA 660
 |||||
 QY 535 HEANKLFKEKEVLYVVSOGELISRLSTKKRYIMKGINNFYKLGOGSKYRVYHNOYSTNS 594
 |||||||
 DB 661 HEANKLFKEKEVLYVVSOGELISRLSTKKRYIMKGINNFYKLGOGSKYRVYHNOYSTNS 720
 |||||||
 QY 595 FALNKHQHRDHKRRHLAKHFCLTTPAGEFKANGSVHGSVLTLSLRLTLTLENNIS 654
 |||||||
 DB 721 FALNKHQHRDHKRRHLAKHFCLTTPAGEFKANGSVHGSVLTLSLRLTLTLENNIS 776
 |||||||
 QY 655 SFLHPNM 661
 :|||:
 DB 777 TSLHPSE 783

RESULT 2
 TRX_DROME STANDARD; PRT: 3726 AA.
 AC P20659; Q27255; Q27327;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax protein.
 GN TRX.
 OS Drosophila melanogaster (Fruit fly).

CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90192757; PubMed=2107543;
 RA Mazo A.M., Huang D.-H., Mazer B.A., David I.B.,
 RT "The trithorax gene, a trans-acting regulator of the bithorax complex
 in Drosophila, encodes a protein with zinc-binding domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
 RN [2]
 RP SEQUENCE FROM N.A.; ALTERNATIVE SPLICING, AND CHARACTERIZATION.
 RX MEDLINE=95009521; PubMed=7924996;
 RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
 RT "The bithorax complex is regulated by trithorax earlier during
 Drosophila embryogenesis than is the Antennapedia complex, correlating
 with a bithorax-like expression pattern of distinct early trithorax
 transcripts.";
 RL Development 120:1907-1917(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Oregon R;
 RX MEDLINE=96100387; PubMed=8555104;
 RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "Conservation of structure and expression of the trithorax gene
 between Drosophila virilis and Drosophila melanogaster.";
 RL Mech. Dev. 53:113-122(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95047388; PubMed=7958911;
 RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "The Drosophila trithorax gene encodes a chromosomal protein and
 directly regulates the region-specific homeotic gene fork head.";
 RL Genes Dev. 8:2478-2490(1994).
 CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
 WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
 CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
 ZINC.
 CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M31617; AAA29025.1; -;
 DR EMBL: 250152; CA90514.1; -;
 DR EMBL: 250152; CA90513.1; -;
 DR EMBL: 231725; CA83516.1; -;
 DR EMBL: 231725; CA83515.1; -;
 DR PIR: A35085; A35085.
 DR HSSP: P20393; 1A6Y.
 DR TRANSFAC: T00850; -;
 DR FlyBase: FBgn0003862; trx.
 DR InterPro: IPR003889; FYRICH_N.
 DR InterPro: IPR003888; FYRICH_C.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001965; ZnF_PHD.
 DR InterPro: IPR001841; ZnF_ring.
 DR Pfam: PF00628; PHD; 3.
 DR Pfam: PF00856; SET; 1.
 DR SMART: SM00542; FYRC; 1.


```

OY 1545 TRDVKYKMKESPPKKTSGTA-----LPSYRK--FWTKSTKKSIFVLPMND 1588
DB 1684 LDSASLGSKQRLTPRKSESEVEDLAGFLFQTPSHTESMNEKTTVSYASQPD 1743
OY 1589 L-----KTLARCGIREVYFYNNAKPAIDIMPYSPRPFEGITMYRLQTVKS 1637
DB 1744 LVDTPSSKPOPKRSLRADTEE-----EFLAFKQTPSAG--KAMHTPKP 1787
OY 1638 LAGVSLMR-LTMASTRMDMAKVPPEGSGSTRFETSETETTEILKRVDVPGIR-- 1694
DB 1788 AVGEKQDINTFLGTPVQKLDQGNLP--GSMRLQTRKKALEBEL-----GEREL 1838
OY 1695 -----FEVCIRKTIIC--PIGV-EPKKEPTTPQKGLRSS-----ALRKPR 1733
DB 1839 FQTPCTDNPADDEKTKKILCKSPQSDPADPHTNKPQKRSKLKADVEEFLFRKLT 1898
OY 1734 -----ETPKQGPVLIETWAAEELELMEIRAF--AEVFE-----KE 1768
DB 1899 SAGRAMHTPKAA-----VGEER--DINTFVGTPVEKLDLGNLPGSKRRPQTPKE 1946
OY 1769 KAAVEQOAKRRLBQOKPTVIATSTSTSPSTSPQAVVAVAPISGVTGTGKAVLT 1828
DB 1947 KAKALEDLAGEKELFQTPGHTEESMTDKITEVCSKSPQDPVATP-----JSSKQRLK 2000
OY 1829 TKVSPATVTEQOKNKHQFPATWVKOGQSNVVOQAVL--GIIPSTGTSGQOTFTS 1886
DB 2001 ISLCK-----VGKFEVLPGVKLQISGKQTOT-- 2028
OY 1887 FQPTATVTVIRPNTSGSGGT-----TSNSOVI-----TPQIRPGMTVIRTP--L 1929
DB 2029 -----HRETAGGCKSKIAFKESAKQMLDPANYGTEMRP-----RTPEEAQSL 2073
OY 1930 QOSTLGAIIITPVWVGAPQOVMTOIRGQPVSTAVSAPNTVSSYPOKSLTSATSTS 1989
DB 2074 EDLAGFELQTPDHTBESTTDDTKTKI-----ACKSPPEESMDTPST-- 2117
OY 1990 NIGSASQPPRPOGOYKLF--MAQLTQLQHGNGNGLTVIQQGQOTTGQLQLIP-- 2044
DB 2118 -----RRRRKTPLGKNDIYELISALKQLQO-----TTHTDKVPQDE 2153
OY 2045 -QGVTVLPGGQOLM-----QAAMPNGTVQ-----RELFL--TPLATTTATAS 2083
DB 2154 DKGINVRETAKOKLDPAASVTSKSKQPRPKKAPLEDLAKELFQTPVCTDKPPTH 2213
OY 2084 TTTTIVSTTA-----ACTGE--OROSKLS--PQMOVHODKTLPPAGSSVGPAPAKQOTA- 2134
DB 2214 EKTTKIACRSPQDPVQTPPTIFKPOSKRSLRKADVEESIALKRPVSQKAMDTPKPA 2273
OY 2135 -----OPSAKPOQOTOPQSPAQPEVQOTQPEVQOTTVSS-----HVP 2171
DB 2274 GDEKDKAKFNGTVOKLDLPGNLPGSKRPQTPKE-----KAAQLEDLAGFKELFQTP 2326
OY 2172 SEAOPTTHAOSKPOVAQOSPOSNVQOSPVVAVOSPQTRIPS----- 2215
DB 2327 GTDKPT--TDEKTKIACKS--PQ-----PDDPYDPASTKQPKRKLKADVEEFLALR 2377
OY 2216 --TPS-----QUSPQGOQOVQTTSQPIPIQPHSL--QIP--SQGQPOSQP----- 2256
DB 2378 KRTPSAGKAMDTPKPAVSEKININTEFTPVQKLDLGNLPGSKRPOQTPKEKALEDL 2437
OY 2257 -----QVQSSQTLQSSQGLINOVSVSSPSRPOLOIQOPQOYAVP-----QIQOQOYV 2305
DB 2438 VGRKELFQTPGHTEESMTDKITEVCSKSPQESKRTSSSKORLKLIPVAKDMEEPILA 2497
OY 2306 LSGI--QSOQVAVQIQAOQSGVPOQIKL--QLPQIQOQSSAVOTHOJONVVT--VQAAVQF 2360
DB 2498 VKRLRTSETQOTHTPEPGDSKSIKAFKESQKQILDPAASVTSGRROLRTKEXARALE 2557
OY 2361 QLORVQOL-----RQOQOK-----KKQOQIEIKR-- 2384
DB 2558 DLYVDEKELFSAGHTEESKTDKNTKIPCKSPPELDTATSTKRCPTKRKEVKEELS 2617

```

```

OY 2385 --EHTLQASNOSELIQOVYAKHNAVIEHLKO--KESMTPAEENORATVQWVKYIL 2440
DB 2618 AVERLTQSGSTYHTEHPASGDEG--IKVLKORAKKPNVVEEBSRR----- 2664
OY 2441 DKIDKEKQAAKKRRESEYQKRSKONATKLSALLFKHKEQORA-----EI 2487
DB 2665 -----RPRAPKEKQPLEDLAGFTLEISHTSHTQESLTAQKATKIPCESPLEY 2713
OY 2488 LKRALDLKDLQLEVOELKRLDKIKKEDLMQLOA----- 2523
DB 2714 VDTYASTKRHLRTYVQK-----VOVKEEPSAVFTQSGETTDADKEPAGEDKIGALKE 2768
OY 2524 -ATVAAKCPVTPVLPAPPAPPSP-----PPGQHTGLSTPLPVAASKKR 2574
DB 2769 SAKOTPAASVTSGRRRPAPRESAQAIEDLAGFQDPAGHTEESMTDKTKIPCKSS 2828
OY 2575 EEEKSSSKKKKKKMISTSKETKDKTYCICQTPYDESKFYIGCDRONMYHRCVGI 2634
DB 2829 PELEDATSSKRRPRPQAKVEYKEE--LLAVGK-----L 2861
OY 2635 LQSEAEILIDYVCPQOSTEDAMTVLPTLEKDYEGIKRYLSLQAKMAMPLEPVDN 2694
DB 2862 TQTSGE-----TTHTDKPEVGEKGTKA-----FKQAPAKN 2892
OY 2695 -DAPDYGVYKEPM-----DLATMBERVQK--RYYEKLETFYAD 2730
DB 2893 VDAEDVIGSKRQPRAPKAKQPLEDLASFQELSCQTPGHTEELANGAD 2940

RESULT 4
AMNH_YEAST
ID AMNH_YEAST STANDARD: PRT: 1367 AA.
AC P08640: P08068:
DT 01-AUG-1998 (rel. 08, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Glucanase 51/52 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STAL OR STAZ OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Cosey T., Dear S., Devlin K., Fraser A.,
RA Goulet S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RC MEDLINE=87194600; Pubmed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STR1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RA MEDLINE=89031230; Pubmed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STZ2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -I- SIMILARITY: TO S.POMBE SPBC215.13.
CC -I- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: Z38061; CAA86176.1; -
CC EMBL: M16164; AAA35014.1; -
CC EMBL: M16165; AAA35015.1; -
CC EMBL: X13657; CAA32069.1; -
CC PIR: B26877; B26877.
CC PIR: A26877; A26877.
CC PIR: S48478; S48478.
CC SGD: S0001458; MUC1.
CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
CC Signal; Multigene family.
CC SIGNAL: 1 21
CC CHAIN 1 22 1367
CC DOMAIN 210 1367
CC CARBOHYD 817 817
CC CARBOHYD 874 874
CC SEQUENCE 1367 AA: 136110 MM: 91C00E2DBD61AA9D CRC64:
CC -----
Query Match
Best Local Similarity 21.2%; Score 401.5; DB 1; Length 1367;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;
QY 1135 DEFG---KACGSE-----SNSTLENSPTVSIQDSSEEDMIVQNSIESIQEFT 1182
DB 191 DFGFYWMNIDCDNNGCGTKSTSTSTSESTSTSTSESS-----TTSTSTSESTP 243
QY 1183 REQDVEVLPLKCELVSGSTGNC--EDRLPVGTEANKCKPSQOKKLEBPVAKCSDQI 1240
DB 244 TSTSTSSSTSTTAPATPTTSTCKEKPPTTSCTEKPP-----PHHDTTPT 296
QY 1241 KLTNTDKNNRESEKKGQRTSTFOINGKONKPKIYLKCELCIESRVSQVNEPK 1300
DB 297 KKTITTSKCTKTTTPVPPSSSTTE---SSAPVPPSSSTTESSAPVTSSTTE-- 350
QY 1301 VNNINKIIPENDIKSLTVEKSAIRPINDVIMEDFNERNSSSTKSHLLSSAEGNYRD 1360
DB 351 ---SSAPVP---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY 1361 SLETLSTKESDS---TOTTPSASCPESNSVQVDEMTESEVKKVTSPTSE---- 1413
DB 395 VPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV-TSTTESSAPVTSSTTES 453
QY 1414 -----EESN--LSNDFIDENGLPINKNNVNGESRKTVITEVTMTST-VA 1457
DB 454 SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV 513
QY 1458 TESKTVIKVEKGDQTVSGSTENCASVTVTVTYKLTSTPSTGSGVDLIISVKEQSKTV 1517
DB 514 TPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSS--STTESSSTP 568
QY 1518 VTTTVTDSLTTGCLVTSMTYSKE--YSTROKVKLAKFRPKTKTSGRALPSYKRFYTK 1575
DB 569 VYSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS 628
QY 1576 STPKSIFVLPNDLKLARKGIRVYFNYNKAPALDIMPVSPPTPGITWRRLQTV 1635
DB 629 STTESSA-----PVTPSSS-----TTE 647
QY 1636 KSLAGVSLMLRLMASLRMDMAKVPVPGGSTRTEETETEIT-----TEIKR 1685
DB 648 SSAPVPTP-----SSSTTESSAPVPTPSSST-TESSAPVTSSTTESSAPVTSSTTE 701
QY 1686 RQVPGIFGIFECIRKIIICPGV-----ETPKETPTQQRKLBSAL----- 1728
DB 702 SSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS 761

```

```

QY 1729 ---RKRRETPKGTGYIETWVAEELWEIRAFERVEKEKAQAVEQAKRLEQ 1784
DB 762 APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSNTSS 821
QY 1785 KP--TYIATSTTS-----PTSSSTTISPAAKVMVAPISGVYTGKVLTKYGS-- 1833
DB 822 APSTPSSSTTESSAPVPTPSSSTTESSS-----APVSSSTTESSAPVPTPSSSN 874
QY 1834 -----PATVTFQGNKNPHOFATWVKGQSGNVQVQKVLG-----IIPS----- 1875
DB 875 ITSSAPSSSTPSSSTTESFST-GTTVTPSSSKYPGQSTETSVSTTEFTIYPTTTSVTT 933
QY 1876 -----STGTSQ--QFTSPQRTAVTIRPNTSGSGGTSNSQVYTPQIIRPGM 1922
DB 934 PSTTTTTTVCSTGTNSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTS 985
QY 1923 TVIRTPLO--OSTLG---KAIIRT-PVMVQPCAPQOVMTQIIRGQVSTAVS-----AP 1970
DB 986 TVCSTGTNSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTS 1045
QY 1971 NTVSTPQOKSLNATSTNSIOSS--ASQPPRQOQOV-KLMAOLTLQFHGNGGLT 2027
DB 1046 YSTSTPQGE-TTFTVTKNIPPTIYLTIAPTPSVTTVFTPTTTTVCSTGTNSA-- 1102
QY 2028 VVIQGGQTTGOLQIPQGVTL-----PGPQQLMQAAMNGTVORFLPTLATTTAS 2083
DB 1103 -----GETTSGGS--PKVTTVTPCSTGTGEXTEA-----TLVYTTA----- 1138
QY 2084 TTTTTVSTTAAGTGEOROKSLSPOMOVHDKTLPPAOSVSGAKKAPQIAPQSPARPOQ 2143
DB 1139 -VTTTVTTSTSTGTNSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTS 1178
QY 2144 TOPSPAPQEPVOTQPEVQOTQTVSSHVPEAOPTHAQSSKPOVAQSQPOSNVQSGSPVR 2203
DB 1179 -----APVTPATNAVPTTITTECSAATNAGETTSVCSATTIYSSAGNTTPSAT--- 1231
QY 2204 VQSPQSTRIRPST--PSQLSPGOQSOVQTV--TSQPIPIQPHTSIQIPSGQPOQSPQV 2258
DB 1232 --TPVTTAT-PTTVITTESSVGTNSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTSAGTTS 1287
QY 2259 QSSQTQLSSGQTLNOXSVSSPSRPQLOIQPOPOVIAVPOLOQOVVLSIQSGVVAQIQ 2318
DB 1288 ATAT-----NPISIKTTS-----QLATPASASVAPV 1315
QY 2319 AQQSGVPOQIKLOPIQIOQSSAVQTHQIONV--VTVOAAS 2357
DB 1316 TSPS-----LTGPIQASGSAVATVSVPSISTYQGA 1348

```

RESULT 5

MUC2_HUMAN

ID MUC2_HUMAN STANDARD; PRT; 5179 AA.

AC 002817; Q14878;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Mucin 2 precursor (Intestinal mucin 2).

GN MUC2 OR SMUC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Intestine;

RA MEDLINE=94132002; PubMed=8300571;

RA Gum J.R., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;

RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.

RT Identification of the amino terminus and overall sequence similarity

RT to prepro-von Willebrand factor."

RT J. Biol. Chem. 269:2440-2446(1994).

RN [2]

QY	1426	NGLPINKNEV-----NGESKRKTIVTEWTMTST---VATESKVIIVKGDQJVV	1475
db	3509	--TPITTTTIVTPPPPGTGTPTTTTITTTTIVTPPPPPGTGTPTTTTITTTTIVTP	3566

FT DOMAIN 2277 2281 CORNR BOX OF ID2.
 FT DOMAIN 58 64 POLY-GLN.
 FT DOMAIN 593 602 POLY-ALA.
 FT DOMAIN 1044 1047 POLY-PRO.
 FT DOMAIN 1713 1718 POLY-ALA.
 FT DOMAIN 1968 1979 POLY-SER.
 FT VAMPPLIC 2333 2371 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1952 1952 I -> T (IN REF. 2).
 FT CONFLICT 2090 2090 A -> P (IN REF. 2).
 SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64.

Query Match 2.7%; Score 382; DB 1; Length 2453;
 Best Local Similarity 18.9%; Pred. No. 1,4e-06;

Matches 434; Conservative 302; Mismatches 782; Indels 782; Gaps 108;

QY 298 HKVGVTDCAVIEIOKNRPYIRH-----EPIGYRSRRKYWFNLRLLEED----- 343
 DB 378 HEISEIIDGLSEQENNEKOMQLSVTPMFMFAEORVKNFTLNMGLMEDDPKAYKDRQFM 437
 QY 344 ---TENENE-----KKIWTYSTRVOLAELIDCLDRDYWEALCKTLEEMREIEIR 390
 DB 438 NWMTDHEKELEFKDKFIOHPKNFGLIASYLEKRSVPCVLYY---LTKNENYKALVRR 493
 QY 391 HMDITEDLTNKRAGSNKSFIAANELLERAKKGDIDVKSPEETEKRNETENDSKD 450
 DB 494 NY-----GRRGRNQOLARPSOEKVEKEEDKA-----EKTEKEEKEKDEEK 538
 QY 451 AEKREEFEDOSLEKDDDKTPDDDDPOGKSEVDFKSEKNGELSPGAGAGSGSTR 510
 DB 539 DDK--EDSKETTKDKDEMTA--EEPEE-REQVTPRKRTAN-----SQRGKGR----- 584
 QY 511 IITPLRPNDSKLSQKQVAAAAAHEANKLFKEGKENVLVNISOGETISRLSTKR-EYIMKG 569
 DB 585 -VTR-----SMTSAAANAANAATEEPPPLPPEPEPTEPEVETSRWTEEEEMEVAKKG 638
 QY 570 NINNYFKLGOEGKRYVYHNOYSTNFPALNKHQREHDHDKRHLAHKCULPAGEFKWGS 629
 DB 639 LVEIGHRWMAIAKMWGCKSEAOKNFEFN-----KRRH----- 672
 QY 630 VHGSKVLITLRLTTQLENNIPSSFLHPNMASHRAMIKAVOMSKPREFALALALE 689
 DB 673 -----NLNLLQO---HKOKASR-----KPRE---ERDVSG 697
 QY 690 CAVKPVVMLPIREPLGHTRLHMTSIEREKRYVKKKEKKEEETMOQATVVKTPPV 749
 DB 698 C-----ESVAST-----VSAQDEDEIEASNEENPEDSGAENSSDTE--SAPS 739
 QY 750 KHOWMKOGKE---YRVTYGGMWSIKTHV-----YRFVPKLPGNTVNVYRKSL 796
 DB 740 PSPVEAKSEDESENMAASRGNTPEVALEATTDPAPCASPSAVP-----TTKPAERESV 795
 QY 797 EG-----TKNNMDENMDESDKRCG--RSPKKIKIEPSEKDEVKGSDAKG-ADQ 844
 DB 796 EAOVTDASAETAEPMQDVHHEEGAGSGSVLDPAPTAKDSVPEMQVEMNTASKGEGDA 855
 QY 845 NEMDI---SKITFKKODY-----KELDSDSDKCKE-----PMVY 879
 DB 856 KENDLSTSEKTFARDEDVVAEQIERPEQSDSDSATCSADEGVGDEPERORVFPMDA 915
 QY 880 DDDMKE-----SH-----VNCQSSQDVVNVSEGFLLPTS 911
 DB 916 KPSILTPPGSILISPIKFNLLDLPLQGHRAVIVPMVSCTCN-IPICTPVSGAL--- 971
 QY 912 YKKKTSKSLDGLERRIKQFTLEEKQRLKLEGGIKIGICKTSTNSKN-----LSRS 966
 DB 972 YORHKIAMHESALLE-----EQROEOYDLCCR--STSPCSTSKSPNEMEVLPDA 1022
 QY 967 P--VITKAKEGQSDSMRQESP-----NANNQDEPDLIOGCSQSDSSVLLKMSDP 1015
 DB 1023 PQGVITINLPEGRALPTTRTPRPPLIBSSKTTVASSEKRSFTMGISISQCGPGTILS--S 1080
 QY 1016 HTTNKLYPKDRLDVDSIRSPETCKPONSIENTNDIEKVSVDLASGOBPTSKTGNF 1075

DB 1081 H--NQAYPO-----EAPKPSVGSISGL-----PROESTKA----- 1110
 QY 1076 IDSKTASADDIGTLICKNKKPLI---QEEBTDIVSSSKSALHSVPKSTNDRTATPLSR 1132
 DB 1111 -----APLTYIKQEEFS-----PRQNSQPEBLVLR 1136
 QY 1133 AMD---PECKLGDSESNSTLLENSSDTVSIQDSSEEDMAYONSNEISIEQFTRDEQVEY 1189
 DB 1137 AQHEGVVGRGAGVQOEGSTIRGTPASKI-----VEET 1168
 QY 1190 LEPKCELVGSESTGNCEDRLPVKG--TEANGKPPSQOKKLEERPYNKCSDOIKIKNT- 1246
 DB 1169 ISSLRSGITNGTFA-----LPAGIPTALVGPVPSRPIESSPEKREELAKGHVY 1222
 QY 1247 -DKNN-----EKRESEKKGQRT--STFOINCQNKPKI--YLKGECKEISESNVVG 1295
 DB 1223 YEGKSGHILSYDNIKANARETRSPRAHEMSLKRSEAEVAGSIKQMSRSPVAPLEG 1282
 QY 1296 NVEPKVNNINKLIPENDISLTPKESAIRPFIINGVIM-----EDNE-----R 1339
 DB 1283 LI---CRALPRGSPHSDLKERTV-----LSGSIQGTPTRALESFEDGLKYPQOLKR 1331
 QY 1340 NSSETKSHLLSSDAEGANYRDSLETLPSTRESDDTQTTTPSASCESNSVNOVEDMEIET 1399
 DB 1332 ESPPIRA--FEGAITKGKPYDGITTI--KEM-----GRSHIHPRODILIT 1372
 QY 1400 SEVKKY-----TSSPTSEBSLSNDFTIDENGLPINKENYNGESKRTVITEVTTMTS 1454
 DB 1373 QESRKTPREVQSTPLI---EGSIS---OCTPL-KPDNNSGGSALIKHNVKSLITGPS 1422
 QY 1455 TVATESKTVIK-----VEKGDKO-----TVVSTSENCAKTV----- 1486
 DB 1423 KLPGRMLEIVPENIKVYERKXEDVKAGEPVARHITSVSSPSVLKSTLHAPPAQOLSP 1482
 QY 1487 -----TTTTTVTKLSTPSTGSAV---DIIVKEOS--KTVVTTVYDLSLTG-- 1530
 DB 1483 GYDSSSARPTVSYONTISRGSPMMNRTSDVSSKSAHERKSTLTPTQRESIPAKSPV 1542
 QY 1531 ---GLVIT-----SMTVSEKYSTRDVKVL---MFSRKKTRSGAL----- 1566
 DB 1543 PGVDPLVSHSPDPHRRSSAGEVYRSHLPTHLDPAMPHRALDPAAYLLQROLSPPG 1602
 QY 1567 -PS-YRKEVTSKTSKSIFFLPNDLKLARKGIREVYFYNNAPK---ALDIMYPSPR 1621
 DB 1603 YPSQYOLVAMENRQTL---NDYITSOQMNYNLK--EDVIRGLSPRRQPLGL-PYAFIR 1656
 QY 1622 PFFGITWKRLOTVAVSLAGVSLMLRLMASLRWDDMAKVP-----GGGSTRTETSE 1674
 DB 1657 GIIDLTL-----NMPTILVPHAGCTSTPPMDRI 1684
 QY 1675 TEITTEILKRDVGPYGIREFEYCIRKILIPGVPEPKETPTQKGLRSSALRPKPE 1734
 DB 1685 TYIPGTQYT-----PPRP-----YNASLSPGHP- 1709
 QY 1735 TPQQTGVIIETWAELELEMEIRAFERYEKEKAQAEQOAKKRLQOKPTVIATSTY 1794
 DB 1710 -----THLA-----AAASAER-EREREREKEREREERERER- 1740
 QY 1795 SPTSSTTSPSAQKVMVAPISGVTYTKMVLTKVGSPTVTFQONKNFHOTFATVWK 1854
 DB 1741 -----ERERIAAPADLYLRPGSE-----QGRPGSHGV----- 1770
 QY 1855 QGOSNGVAVQVQKVLGIIIPS-----STGTSQOTFTSQOPRATVTIRPNTSGSGGITSNS 1910
 DB 1771 --RSSPSVYKROEIIILQORPSVFOGTNGTS--VITPLDP--TAQIRIMPLPSG----- 1917
 QY 1911 QVITGPOIRPGMTYIRPPLQOSTLGKAIIRTPVWVQEGAPQVWTTIRGOPYSTAVSAP 1970
 DB 1818 ----GPSISQGLPASR---YNTAADALAL-----YDAASAP 1948
 QY 1971 NTVSSTGQKLSATSTSNISQASASOPPRPOGQVYLTMAQLTOLTQGHGNGOGLTVI 2030
 DB 1849 QMDVSKTKESKHEAARLEENLRSRAAVSBOOLOEQKNLVE-----KRSVQVCV 1898

DR EMBL; AF026245; AAB86384.1; -
DR EMBL; AF083037; AAD22767.1; -
DR EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR EMBL; AF091711; AAD39719.1; -
DR EMBL; AB018346; BAA34523.1; -
DR EMBL; AC000066; AAC60380.1; ALT_FRAME.
DR Genew; HGNC:379; AKAP9.
DR MIM; 604001; -
KW Colled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-Rit SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 914 COLLED COIL (POTENTIAL).
FT DOMAIN 944 1022 COLLED COIL (POTENTIAL).
FT DOMAIN 1100 1185 COLLED COIL (POTENTIAL).
FT DOMAIN 1253 1280 COLLED COIL (POTENTIAL).
FT DOMAIN 1336 1392 COLLED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COLLED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COLLED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COLLED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COLLED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COLLED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COLLED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COLLED COIL (POTENTIAL).
FT DOMAIN 3587 3689 COLLED COIL (POTENTIAL).
FT DOMAIN 3726 3730 COLLED COIL (POTENTIAL).
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 GLU-RICH.
FT VARSPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 1637 1642 OLOEEL -> LATRD (IN ISOFORM 4).
FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 4).
FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 3).
FT VARSPLIC 2175 2183 SADFQKVE -> Q (IN ISOFORM 6).
FT VARSPLIC 2895 2907 VEGFMYCFSTLC -> GSSIELAHSDAYQREICSS
(IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).
FT VARSPLIC 2895 2948 MISSING (IN ISOFORM 5).
FT VARSPLIC 3901 3911 STTQFHAGMRR -> ALSLTTHQHSAPTAPLPEILLSH
SIG (IN ISOFORM 6).
FT VARSPLIC 1347 1347 K -> KQ.
FT VARSPLIC 1347 1347 /FTID=VAR_010926.
FT CONFLICT 76 76 E -> Q (IN REF. 3).
FT CONFLICT 475 475 M -> I (IN REF. 3).
FT CONFLICT 554 554 E -> G (IN REF. 3).
FT CONFLICT 638 638 R -> S (IN REF. 3).
FT CONFLICT 663 663 N -> S (IN REF. 3).
FT CONFLICT 913 913 H -> N (IN REF. 3).
FT CONFLICT 956 956 K -> N (IN REF. 3).
FT CONFLICT 980 982 QKH -> PKP (IN REF. 1 AND 2).
FT CONFLICT 997 997 Q -> P (IN REF. 1 AND 2).
FT CONFLICT 1001 1001 N -> P (IN REF. 1 AND 2).
FT CONFLICT 1020 1020 N -> D (IN REF. 3).
FT CONFLICT 1028 1028 V -> E (IN REF. 3).
FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).
FT CONFLICT 1703 1703 N -> T (IN REF. 3).
FT CONFLICT 1707 1707 V -> G (IN REF. 3).
FT CONFLICT 1802 1803 MISSING (IN REF. 5).
FT CONFLICT 1843 1843 A -> P (IN REF. 3).
FT CONFLICT 1956 1956 I -> V (IN REF. 3).
FT CONFLICT 2027 2027 V -> D (IN REF. 5).
FT CONFLICT 2157 2158 EI -> HE (IN REF. 7).
FT CONFLICT 2169 2169 E -> V (IN REF. 3).
FT CONFLICT 2514 2514 L -> R (IN REF. 3).
FT CONFLICT 2851 2851 I -> N (IN REF. 8).
FT CONFLICT 2957 2957 E -> D (IN REF. 3).
FT CONFLICT 2963 2963 P -> S (IN REF. 3).
FT CONFLICT 3087 3087 Q -> H (IN REF. 3).
FT CONFLICT 3218 3218 Q -> H (IN REF. 3).
FT CONFLICT 3307 3309 ESE -> OSO (IN REF. 3).
FT CONFLICT 3751 3751 P -> A (IN REF. 3).
FT CONFLICT 3833 3833 T -> S (IN REF. 3).
SQ SEQUENCE 3911 AA; 453664 MM; 3FBICBIC819B47AA CRC64;

Query Match 2.7%; Score 380.5; DB 1; Length 3911;
Best Local Similarity 17.2%; Pred. No. 2.7e-06;

Matches 483; Conservative 475; Mismatches 1035; Indels 821; Gaps 115;
QY 305 DCAVET-----OKNP--YIRHEPIGYDRSRKRY-----FLNRLLIE 341
DB 125 DCSSEVNGCSFVARTGPTMLIREEEGVDDSYEGQAQDSPHLMMESELAKQHEIE 184
QY 342 EDTEENNEKKIWTYSTRKYOLAEILIDCL-----DDYMEAEIKTL 381
DB 185 ELKRELEEMRVTYTGEBLOLOEPEEALIKOROGITOLTANTAOARREDETFRELELT 244
QY 382 EEMRE-EIH-RHMDITEDLTNRKAGSNKSFLLAANEI-----LESTIRAKKG 426
DB 245 EQSQKLIQIOFOQOQASETLRNSTHSTADLLQAKQIILHQOQLEPDHLEDOKKKE 304
QY 427 DI-----DNKSEETFEKDNTEENDSKAENRE-----EFDQSLKDS 467
DB 305 DFTMQISFLOEKIR-VYEMHODK-KVENSKEIOEKETIIEBLMTKIIIEEKKTLEILD 362
QY 468 DDKTPD-----DDEQCKSEVGDFKSEKNGELSEPGAG--KGASGSTRILRLNRP 518
DB 363 KLTPADKLGLQEOLOVQAKNOEIKNNKLELINSKOKEROSSEIKQMGTEIELQRNKK 422
QY 519 DSKL-----SOLKSQVAAAHAHA-NKLEKGEKVLVNSQGEISLSTKEVT 566
DB 423 DSQFETDIQVMBQFQRLKQELRAELDEMGOQIYQMKQELIRQHMQMMEKTRHNGE 482
QY 567 MGNINNYKRIQOEKRYRYHNOYSTNSFALNK--HQHEDDKRRHLAKFCCLTPAGEF 624
DB 483 MENALSYNI-----TYNEQIKLMVAINELNIKIDDTSOKEKEKEEL----- 528
QY 625 KMGSVHSGKVLITSLRLITOLENNIPSPFLHPMAASHRANWIKAYOMGSKPREPALA 684
DB 529 ---GLLEKCALQRLQEDLVEEL-----SPSRD--IQRAQIADQ-SKLEAKKS 576
QY 685 LALICAVAPVYMLPIWREFLGHTRLRHTS---IEREKEKVKKKKQOEETMQQA 740
DB 577 LSTVEDLKAELIVSASESRRELELKHAEVYTNKIKLEMEKKNVLDPRMESQAEIER 636
QY 741 TWVKTFEPYKHQVWQKGE---EYRVTVGGWSMLSKTHYRVFKLQGNINVRKSL 797
DB 637 LRTOLLFSHEELSKLELEIETHRIN-----TEKLNDLNGIHKQOID 680
QY 798 GTRKNNDENMD--ESDKRCSNSPKIKIIEPDEKDEKVSDDAKAGD-----QNM 847
DB 681 GLQNEQSKQIEMQEKONLIRKQOLILIELSKLDQOSLVNSKSEEMTQINELQEI 740
QY 848 DISKITEKK---DDVVEL-----LDSQDKPCKEEPMEV 879
DB 741 ELTROEKEKGTLEBOVEDQLKTELLEKOKKEKENDLOEKFAQLEAE-NSILDEKKTLL 799
QY 880 DDDMKTESHVNOQESQYDVVNVSSGFLHRTSYKKTKTSSKLDGLLEKRIKQFLLEKOR 939
DB 800 EDMKLIHPVSGEE-----RIFLDSTIKSKSDSVWEKEI-ELIIEENED 843
QY 940 LEK--IKLEGGIKIGKSTNSSKNL-----SESPVITKAKEGOSQDMRQ--Q 985
DB 844 LKQOCIOLENEIEKQRFNFSAEKNFEVNYQELQEVYCLLKVADDLDESKNNKELEYKS 903
QY 986 SPNANNDPEDIQGCSSDSSVLRMSDPSTHTNKLYPKDRVLDVDSIRSEP----- 1038
DB 904 KIKALINEB-----LHQRINPTVTKM--KSSVPEDEKTFVAETLEMEGV 946
QY 1039 -----KCPRONSIENDIEKYSVDLASROEPTKSTKTKNDFFIDSKLASADDI 1087
DB 947 EKDITLMEKEIEVTKRELE--LSQRUSDLS---EQKQKHGEISFLNEEVSQKQ-- 998
QY 1088 GTLLCKNNKPLIOESPTIYSSKSAHSSVSKSTNDRDAPTLSRAMDFGKILGCDSESN 1147
DB 999 -----KEQVSLRCHELEIITIHNRH-----ENVSQDITQVS 1029
QY 1148 STLENSPTYSIQSSSEDMAYVONSNSISQFTRQDYVLEPLKCLVSGESTGNCE 1207
DB 1030 STL---DGVVTMTSRGAEGSVSKVNSFGEESKINVEDKVSFE---NMTVGEESKQEQ 1081

RL Gene 105.185-195(1991),
 CC -1- FUNCTION: BINDS TO POLYTENE CHROMOSOMES. SEEMS TO INTERACT WITH
 CC PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER
 CC COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND
 CC SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF
 CC OTHER PAIR-RULE GENES SUCH AS EVE, FTZ, AND H.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: SALIVARY GLANDS.
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC -1- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on lists
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/1sb.ch>).
 CC or send an email to license@1sb.ch).

DR	EMBL; X63672; CAA5211.1; -. COS.
DR	EMBL; M64750; -; NOT_ANNOTATED_COS.
DR	PIR; S23632; S23632.
DR	Flybase; Fgmn004861; pb-p.
DR	InterPro; IPR001660; SAM.
DR	Pfam; PF00536; SAM; 1.
DR	SMART; SM00454; SAM; 1.
DR	PROSITE; PS50105; SAM_DOMAIN; 1.
KW	Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT	ZN_FING 1365 1387
FT	DOMAIN 1513 1577
FT	GLN-GLN 74 80
FT	DOMAIN 411 450
FT	GLN-RICH 494 520
FT	DOMAIN 619 650
FT	GLN-RICH 775 960
FT	DOMAIN 1233 1290
FT	SER/THR-RICH 1234 254
FT	CONFLICT 254 254
FT	MISSED IN REF. 2).
FT	CONFIDENCE 1415 1415
QO	SEQUENCE 1589 AA; 167297 MW; A6DF0CF9106E1891 CRC64;

Query Match	2.6%;	Score 379.5;	DB 1;	Length 1589;
Best Local Similarity	21.3%;	Pred. No. 1.e-06;		
Matches 242;	Conservative 137;	Mismatches 407;	Indels 349;	Gaps 42

QY	1666	RDVGYGIRFEYCIKRLIICPLGVPETPKETPTPOKGLSSALRPKRPTKOTGPVIE	1745
Db	143	HGPTPTGRR-----OTHTP-----STPRRBPSTPNTNCSI--	175
QY	1746	TWVAEELELEIRAPAEVREKEKAQAVEDQAKKRLQOKPVIATSTTSP	1797
Db	176	-----ARHTSLLEKAQPGQ-----VAATTVPRQISPEQLQ	209
QY	1798	---SSTSTISPAQKVVAVDISGVTGTRKMYLTTKVGSPATVTPQOKNHEQTATVWK	1854
Db	210	QEVASNPYALIQWKEPFTHTSS---GTELKATYI-----MEVQQLQLOQLQSLSEANG	261
QY	1855	QGSQNSGCVQVQOKVLGILIPSSGTGSQQTFTSPQRTATVILRP-----NTSGSGGTSSN	1909
Db	262	GGAASAGAGAAS-----PANSQOSQOQ-----QNSTASTSPMQLAATIGVQGDWTO	311
QY	1910	SQ-----VITGQIRPGMTVILRPLOOSTIGKALIRPMVYORCAPQV	1953
Db	312	GRIVQLMQPSTSEFLPQMTVIGMLLHPG-GLQOQPIQVITACKPF-----QGNGPQML	363
QY	1954	MT-----QILIRQ-----PYSTVASAPNVTSSTPGOKSLTSATST	1988
Db	364	TTTTQAKKQMIIGGQAGFAGCNATATCTIPTNHNSPQTVLLESPNNVTSPOQOONLLQS---	419

QY	1989	SNTSSASGPPRRPGOGVKKLMAQLTQTGHSGHNGSLIVVIGGG--GTTCGL-OLIRPG	2046
Db	420	-----MAAAAGAGDGLTQQDQOOFNQDQDQDGLTQ-----QDQDLFTALKAKVGVDQAOKLAQVYK	473
QY	2047	VTVLVRPGGOLLMOAMPNGT-----VDRFLPTPLATATATASTTTVTSTTAAGCE	2098
Db	474	VTTT-----SSAVQATGPGSITGSITQGVQDVQDQDQDQTQTQTQOCVVQSOSTLPVGVGG	529
QY	2099	QKOSKLSPOMOVHODKITLPPAOSSSVGRPAKQ--POTAQPSARQP-----OTOP--	2146
Db	530	Q-----SVQTALLMGAQDAQMDIPIWFLNNAAGLDPEFGNOITILRNPDG	574
QY	2147	-----OSPAPREVOTQPE--VQ-----TOTTVSSHPRSEAQPTHHQSSKPRYVAQSPQS	2194
Db	575	TGGMTIQDQDPATQTLDTQDQNDQIDICNVQDTPTKAKPTDLALPKQDQDQDQDQGITTNQTDQ	634
QY	2195	NVOGSPVRYVQSPSQTRI-----RPSRP-----SOLSPGOQGVQTTTWSPI-----	2236
Db	635	QQLAVALTAQLQDQDQDQQLTAAALDHPGARVPMPHNGTGTVRRASSVSITQDNQSLKAKMRN	694
QY	2237	--PIQPHTSIQDISQGPQSPQPV-----QSTHQ-----	2263
Db	695	KQDPRPALATLKTEIQGVAGKNKVGGHLTTVYQDQDQATNLQDVNNAAGNKVMYSTGT	754
QY	2264	--TISSGQTLNQVSSSPSRPOLQIDOPQPVAVFQLQDQGVVLQSIQSOQVNAVQIAQ	2321
Db	755	PITLNQGTTHAAATAAGVDKQDQQLDFQKQDI-----LQQQ-----QMIDQDIALDQWQ	805
QY	2322	SCVPQQLKQLPDIQOSSAVQTHQIQNVTVQAASVDEQLQRVQLRDQDKKKKQDIIE	2381
Db	806	Q-----QAAYADQDQDQDQDQVQSDQDQVNAQDQDAVDAQDQDAVDAQDQDQDQDQVQAQ	857
QY	2382	IKREHTLOASNQSEIIQKVYMKHNHVIEHKQKSMTPAEENEORNTVQVWKIYD	2441
Db	858	AQAQHQAALANATQILLQ--VANQPFITSH-----QDQDQDQLNHQLLIQDQLQ	903
QY	2442	KIDKEEQAOKKKRKREEVEQKRSKANATKLSALLEFKHEQDLRAELIKRALLEDKDQIE	2501
Db	904	QQAQAVQAVQAVQAVQADQDQDQDQDQEQDQONIFIQ-----QIV	966
QY	2502	VOEFELKDKLIKKEKDLMLQAQNTVAACPCPVTVLPA-----	2540
Db	937	VQDSGATSQOTSQDQDQHMSGQLQLSSVFPVSSSTTPAGITSALQALISAGALFQT	996
QY	2541	--PAPPSPPPPPEGVQHTGILSTPTLPVASOKRRKREEKDSKSKKKMISTT	2593
Db	997	AKPCTCSSSPFSSTSVYTITMOSSTPLVTSTTVAISQQAQQTQSAQVHQHQQLISAT	1051

RESULT 9

ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
ID ANK2_HUMAN	001485;	
AC AC	001484;	
DT DT	01-APR-1993 (Rel. 25, Created)	
DT DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE DE	Ankyrin 2 (Brain ankyrin) (Ankryfin B) (Ankyrin, noneerythroid).	
CS CS	Homo sapiens (Human).	
OC OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
NC NC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX OX	NCBI_TaxId=9606;	
RN RN	[1]	
RP RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).	
RC RC	TISSUE=Brain stem;	
RX RX	MEDLINE=91302466; PubMed=1830053;	
RA RA	Oto E., Kunimoto M., McLaughlin T., Bennett V.;	
RT RT	"Isolation and characterization of cDNAs encoding human brain	
RL RL	ankyrins reveal a family of alternatively spliced genes".	
J J	Cell Biol. 114:241-253(1991).	
RN RN	[2]	
REP REP	Carpenier S.;	

RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RC TISSUE-BRAIN stem;
 RX MEDLINE=94075409; PubMed=8253844;
 RA Chan W., Kordeli E., Bennett V.;
 RT "440-kD ankyrinB: structure of the major developmentally regulated
 domain and selective localization in unmyelinated axons."
 RL J. Cell Biol. 123:1463-1473(1993).
 [4]
 RP SEQUENCE OF 463-495 FROM N.A.
 RX MEDLINE=92009921; PubMed=1833308;
 RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
 Lux S.E., Ward D.C., Forget B.G.;
 RT "Isolation and chromosomal localization of a novel nonerythroid
 ankyrin gene."
 RL Genomics 10:858-866(1991).
 CC -1 FUNCTION: Attach integral membrane proteins to cytoskeletal
 elements. Also bind to cytoskeletal proteins.
 CC -1 ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1 TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
 CC CELLS THROUGHOUT THE BRAIN.
 CC -1 PTM: PHOSPHORYLATION AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
 AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
 AND FUNCTION (POTENTIAL).
 CC -1 SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC -1 SIMILARITY: CONTAINS 1 DEATH DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 CC EMBL: X56957; CAA40278.1; -
 DR EMBL: X56958; CAA40279.2; -
 DR EMBL: 226634; CAB42644.1; -
 DR EMBL: M37123; AAA62828.1; -
 DR PIR: S14533; S14533.
 DR PIR: A39643; A39643.
 DR PIR: B39643; B39643.
 DR PIR: S14569; S14569.
 DR HSP: P42771; IDC2.
 DR Genew: HGNC:493; ANK2.
 DR MIM: 106410; -
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000488; Death.
 DR InterPro: IPR000906; ZUS.
 DR Pfam: PF00023; ank; 24.
 DR Pfam: PF00531; death; 1.
 DR Pfam: PF00791; ZUS; 1.
 DR PRINTS: PRO1415; ANKYRIN.
 DR SMART: SM00248; ANK; 21.
 DR SMART: SM00218; DEATH; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 20.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
 phosphorylation.
 FT REPEAT 63 92 ANK 1.
 FT REPEAT 96 125 ANK 2.
 FT REPEAT 129 158 ANK 3.
 FT REPEAT 162 191 ANK 4.
 FT REPEAT 193 220 ANK 5.
 FT REPEAT 232 261 ANK 6.
 FT REPEAT 265 294 ANK 7.
 FT REPEAT 298 327 ANK 8.
 FT REPEAT 331 360 ANK 9.
 FT REPEAT 364 393 ANK 10.

FT REPEAT 397 426 ANK 11.
 FT REPEAT 430 459 ANK 12.
 FT REPEAT 463 492 ANK 13.
 FT REPEAT 496 525 ANK 14.
 FT REPEAT 529 558 ANK 15.
 FT REPEAT 562 591 ANK 16.
 FT REPEAT 595 624 ANK 17.
 FT REPEAT 628 657 ANK 18.
 FT REPEAT 661 690 ANK 19.
 FT REPEAT 694 723 ANK 20.
 FT REPEAT 727 756 ANK 21.
 FT REPEAT 760 789 ANK 22.
 FT REPEAT 793 822 ANK 23.
 FT DOMAIN 1773 1950 REPEAT-RICH REGION.
 FT REPEAT 1773 1784 REPEAT A.
 FT REPEAT 1785 1796 REPEAT A.
 FT REPEAT 1797 1808 REPEAT A.
 FT REPEAT 1809 1820 REPEAT A.
 FT REPEAT 1821 1832 REPEAT A.
 FT REPEAT 1833 1844 REPEAT A.
 FT REPEAT 1845 1856 REPEAT A.
 FT REPEAT 1857 1867 REPEAT A.
 FT REPEAT 1868 1879 REPEAT A.
 FT REPEAT 1880 1891 REPEAT A.
 FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
 FT REPEAT 1903 1914 REPEAT A.
 FT REPEAT 1915 1926 REPEAT A.
 FT REPEAT 1927 1938 REPEAT A.
 FT REPEAT 1939 1950 REPEAT A.
 FT DOMAIN 3536 3620 DEATH.
 FT VARSLIC 1039 1039 Q -> QFGKHLPLAPPLNAGESLVSRIIDGPRGK
 (IN ISOFORM 2).
 FT VARSLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT CONFLICT 475 476 GO -> PE (IN REF. 4).
 FT CONFLICT 971 971 I -> S (IN REF. 1).
 FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).
 FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
 SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;
 Query Match 2.6%; Score 375.5; DB 1; Length 3924;
 Best local Similarity 17.7%; Pred. No. 4,1e-06;
 Matches 550; Conservative 398; Mismatches 1121; Indels 1037; Gaps 126;
 QY 16 TQDSEDEDEDEDDDDSDYPREMEDDDDASYCTESSFSRSHYSTSSTGRKRRPVYHRP 75
 DB 1423 TKSESDDQEEED-----MTSEKNDTESTETSVLKSILVNEVVLASPLD--- 1470
 QY 76 RPILEEK-----DIPPLEPPSSSEDLMPNPHINAVIAYEVLN 116
 DB 1471 LSEVSEKODLIMTALITLTVSDKAGSIRKVLVAABE--BGEFFELVERKEDLEK 1528
 QY 117 FGVLRSPREFDFCAALVSOCSTLMAEMHYVLLKAVLREDDTSNTPRGPADL----- 171
 DB 1529 VNEILR-SCICTDSESSVSSRSERGLVEEYVI-----VSEDEIEAROKAPLEITIEYPC 1583
 QY 172 -----KDSVNSTLYFDGMTV-----PEVLRYVCESD 198
 DB 1584 VEYRIDKEIKGVKEDSTGLVNLITLDDLNTCPVLRKQLOTYODKAGKKEALAVGSSSE 1643
 QY 199 KEYHNALP-----YQEAEDRYG---PVENKIKVLQFLVDQFLTINIARELLMSGVYQY 250
 DB 1644 KEKQDIPRDETOSTOKNPKPSLGIKRPVRRKLEKQKQKEEGIQASAEKAL----- 1695
 QY 251 DDICRGVCHKGLDLCSETGSAVYHNLBCVRYPLEYVPEDEMCVCAHNVYGVGT-DCVAE 309
 DB 1696 -----KKGS-----SESLSD-----PGLAPPLRP 1717
 QY 310 IQKNKYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENEKKIWYSTVKQLAELIDCID 369
 DB 1718 VKATSPLIETETIGSIKDKYK--ALOKRV-----EDQK----- 1749
 QY 370 KQYWEALCKILEMRBEIHRHMDITEDLTNKRAGSNKSFLLAANBELLESIRAKKGDDID 429

Db 1750 -----GRSKLPVRKGEDEVPKRT--THRHPDASPS--LKSERHAPG--- 1788
QY 430 NVKSEETKDNKNETENDSKDAEKREFEEDQSLKOS-----DDKTPDDDEQCKSEVGD 485
Db 1789 ---SP-SKTERHSTLSSAKTERHPVSPSSKTEKHSVPVSAKTERHSPASSSS----- 1840
QY 486 FKSEKSNELSESPQACGASGSTRITIRLRNPNDSKLSQOLKSOQVAAAHANKIFKEGK 545
Db 1841 -KTEK-----HSP-----VSPSTK--TERHSPVSTTEHHPVSPSG-----KTDK 1879
QY 546 EVLVNNGEISRLSTKKEVIMKGNINNYFKLGQCKRYVHNQYSTNSPALNKHQRED 605
Db 1880 RPPV-----SPSGRTEKHPVSPGTERKRLPVSPSGR-----TDKHQPVST 1920
QY 606 HDK-RRHLAHKCTLPAGEFFKNGSVHGSKVLTJSTLRITTOLENNIPSSFLLHPNASH 664
Db 1921 AGTEKHLP-----VSPSGTEKOPVSP--TSKTERIETTMVREL----- 1960
QY 665 RANNIKAVOMCSKPRERPALALILECAVPVYMLPIWREFLGHTRLHRMTSIEREEKY 724
Db 1961 ---MKAFQSGODPSKHKTGLFEHKSQK-----QOQEKGV 1994
QY 725 K-KKEK-----KOEHEETMOATVWKTTPYKHQVWKQGEVRYTGYGKSWTSTHY 778
Db 1995 RVEKEGPILTQREAOQTEQNTI-----KROQRLEPVTG----- 2027
QY 779 RVEKRLPGNTVWYRKSLGKTKNNMDENESDKRCSRPKIKIPDSEKDEVKSDA 838
Db 2028 -----TAESKRGVRSISGVKK--EDAAGKEKVLSHKIRPEVQSVPEHSHRESVPR 2079
QY 839 AKGAD-QNEMDISKITEKK-DQVYKELLDSDSKPCKEEPMEDDDN-KTESHVNOCESS 895
Db 2080 EKMADEGDDMDLOISPRKSTJDFSEVYIKQELBDNDKYOQFRLSETEKQALHLD----- 2134
QY 896 QVDVYVNSEGFHLTSTYKK-----KTGSKLDGILE-----RRIKOP 932
Db 2135 QVLTSPNTEFPL--DYMKDEFLPALSLQGALDGSSESLKNGVAGSPGCSLMEGTPOI 2192
QY 933 TLEBKOLKLEKIGKIGKIGKSTYNSKNLSBPVYTKAKEGCOSDSMRBOESPANND 992
Db 2193 SSESSEYKHE-----GLAETPETSPELSFSPKSEQGETKSTETETTELRESE 2243
QY 993 Q-----PRDLQOGCOSQSSSVLRKMSDPSHTNKLXPKDRLVDVYINSPEKCKKONSTEN 1048
Db 2244 KEHPHTKIDITGSEBERKAV--TEDESETSESFOKEATLG-----SPKDTSPKR--OD 2292
QY 1049 D-----LEKVSIDLASR--GOEPTKSKTKGNDPFLIDSKLASADIGI 1089
Db 2293 DCGSGSVALAKETPTGLTEAACDDEGQRTFGSSAHKTOT-----DSEAOES--T 2340
QY 1090 LITKNNKPLIOESDITVSSSKSALHSVPRSTNDRDAPPLSRAMDFEGKLGCDSESNST 1149
Db 2341 ATSEDEKALPLPRAVSKYKTDGTGTESKPOGVIRSPQGLELALPSR-----DSEVLAS 2390
QY 1150 LENSSTVTSIODESSEDMIVONNESISQGFRTREDDVLEPLKELYSGESTGCCEOR 1209
Db 2391 VADDSILAVSHKDSLEASPVLEDNSS-----HKTPDSELP-----SPIKESPCRDS 2435
QY 1210 L-----PVKGTENGKPRS-----OQKLEERPN 1234
Db 2436 LESSPVEPKMKAGIIPSHPLPAVAKTELLETVASVRSLRLDPOGSAEDDSLEBTSLM 2495
QY 1235 KCSDOJLKLANTYDKNNENRESEKKGORTSTQJINGKDNKPKIYKLGCEKLETISESRVVS 1294
Db 2496 ESSGKSPSLPDPSPSESEVYEYTPKTTDVST-----PKPAVTHECAEED-----S 2541
QY 1295 GANEPRVNNINKIIPENDIKSLTVKESAIRPFLINGDVIMEDNENRNSSTKSHLSSSDA 1354
Db 2542 ENGEKK-----RTPPEEFKMYTKIKMPDELEQAKQKRDYKKEKQESS--SSDP 2593
QY 1355 EGNRYDLETLPLSTKESDSTQTTTSPASCPESENVSQVDEMELETSEVKKVTS----- 1407
Db 2594 DADCSVDVDE-FKHNGSGDESGVPLVYTESKRVSSSESEPELALQKKAGDSGLPPE 2652

QY 1408 -----SPITSEESNLSDFT-----DENGILPKNNENNGES----- 1440
Db 2653 VIRQVPPSLSSDNSSSPEEVOFQPVVSKOYTFKKNEDTOEERPKSEEEKSESHLAE 2712
QY 1441 KRKTIVTEVMTSTVATESKTVLKVKGDKQVAVSTENKCASTVTTTTTKLSTPS 1500
Db 2713 DRHAVSTE-----AEDRSYDKLRNDQKRICDHGCCAMSPSSARPPVSGIQSP 2763
QY 1501 TGGSVDIILSVKEOS--KTVVTTTTVSDSLTTGTGLVTSMT-----VSKESTRDKV 1550
Db 2764 TGDVD-----EDPVYKESIALQGTHEKDTGEBELDVSAESPQADCPSESPSSSLP 2818
QY 1551 LMKFSRKRKTRSG-TALPSRKVEYTKSTKSTIVLPBDLKLARKGGLREVPYVYNAK 1609
Db 2819 HCLVSEKLEDEDISATSSIOKTEVTKTDETFENLPKD----- 2856
QY 1610 PALDIWYPPRPPLTFTGTYWYRLQTVKSLAGVSLMLRLWASLRMDMAKVPGGGSTR 1669
Db 2857 -----CPSQDSST--TOT-----DRFSMDVVSOLAEN 2883
QY 1670 TETSEETITTEIIRRDVGPVIREYCIRKILCPGP-----ETPKETPPOR 1721
Db 2884 DETYDPOITS-----PYE-----NVPQSFSSSESKTQTDANHTT 2919
QY 1722 GLRSSALRPKRPEP-----KOTGPVILETWAELELELMEIRAFARVEKEKAQAV 1773
Db 2920 SFHSSEYVSTTSPVEDVVASSSSGTVLSK--ESNEFGODIK-----M 2962
QY 1774 EDOAKRRL-EQOKPVIATSTSTPTSTSTISPAQVWAPISGSVTTGTMKVLT--TRY 1831
Db 2963 ESQLESTLMEWQSDSV--SSSEPTMSATTTVGEQI-----SKVITIKTV 3007
QY 1832 GSPATYTFQONKNEHOTFATWVQOGSNGSVVOOVKULGIIPBSGTSGQFTSFQPT 1891
Db 3008 DSD--SWSEIRDEDAFARVYK-----EEQKIRGLM-----YDRO----- 3041
QY 1892 ATVTIRPNTSGSGGTTNSQVITGPIRPGMTVIRPLOSILGKALITPVMVPGAPQ 1951
Db 3042 -----SGGT-----PDITPARKIPEGT--PISQONPLFBEKG-- 3074
QY 1952 QVMTQIIRQOPVSTAVSAPNTVS--STPGOKSLTATSTSNIOSASOPRPOOGVYL 2008
Db 3075 --LEFMTRSQALIDMTKRSYADESFHFQIQGSEBRETLSEDEVKAGTADPLPLE----- 3127
QY 2009 TMAQLQLQGHGNGGLVVIQOGQTTGOLQILPQYV--VLPGGOQLMOAMPN 2064
Db 3128 TSAESIALS-----ESKETVDEADLLPDSVSEVEEELIPASDAQLN----- 3168
QY 2065 GIVQRFLETPLATYATTAATTTTSTTAAGTGEORQSKLSPQMVHODKTLPPAQSSSV 2124
Db 3169 -----SOMGISASTETPTKEAVSVGI-----KDLPTVQTDGI 3200
QY 2125 GPKAKPOPTAP-SARPOPOTO-----POSAPOP--EVQTOBEV 2160
Db 3201 PRLSGVKQIJCPSDSEPAVOQLDFSTLRSYVSDRGDSDPSSEQKSVLEIPTAME 3260
QY 2161 QTOQTWVS-SHVSEAOPT-----HAOSKRP----- 2184
Db 3261 NVPFTESKSKITVRMPITSTAPRPSAEVSESVSEDFLSSVDEENKADAKPKSLPVKP 3320
QY 2185 -QVAAOSOPQSNVQOSPPVWQSPQOTRIRPSTPSQLSPGOOSQVQTTSPITQIPTS 2243
Db 3321 IQRVBQQLSDLTQVQKTYAPQGOUMASIPDNRK--SESDASSLDSKTKCPVTRSTTE 3379
QY 2244 LOIPSGOPO-----SOPVQOSTQTLSSQTLNIV-SVSPSPRPOLO----- 2285
Db 3380 TETESREARELELESEBEGATRPKILTSRLPVKSRSTTSSCRGTSPTKESKEHFDLYR 3439
QY 2286 -----IQQPPQVYAVPQLOQOVVLSQIQSQVVAQIQAOQSSGVPQOIKILOLPQI 2336
Db 3440 NSIEFFEELIDEASKLVRLTQSEREQEIVSDDESSALAEVSVIENLPVETEHSEVPEDI 3499

QY 2337 QOASVQHOQIONVT-----VOASVQBOLORVQO-----LRDOQ 2372
 Db 3500 FDRIMDESJETLIERIPDENGHDHADPDQDERIERLAIADHUGFSTELARELD 3559
 QY 2373 QKKKO-QOIEIKREHTLOASNO-----SEIIQOVVKKHNAV 2408
 Db 3560 FTEEDHRIIRIENPISLDQSYLLKIMLERDGKATDNLVECLTKIKRMIDVILMETN 3619
 QY 2409 IEHLKQKSMTPAREERNORMIVCNQVKKYILD-----KIDKEKQAKRRKRESEVEOK 2463
 Db 3620 TEPLQERISHSVAYAEIQ-----ITLDHSEGSFVLQDEELCTQHKQKEQAVSK 3668
 QY 2464 RS-----KQNTKLSALLF--KHKEQLRAELKKRALLD 2495
 Db 3669 ESETCDHPPIVSEEDISVGYSTFDQGVKRTGESSSTALLFPQTHKEOVQODPSGKKMODLP 3728
 QY 2496 KDQLEVOEELKRDCKIKKENDLMQQAATAVAPCPVTP-----VLBPAPAPPPSP 2550
 Db 3729 EESSLEYQOEY-----FVTTPTGETSETQKAMIVPSSPSKTPPEVS 3769
 QY 2551 PPGVOHTGLLSTPT-----LPVASQKRRREBEKDSKSKKKMISTTSKTKKDKTKLY 2604
 Db 3770 TPAREEKLYLQPTSSERKSGSPITIQEPPEPSHRESS-PRKTSLVIVESADNQPET--- 3825
 QY 2605 CICKTPYDESKFYICDCQNTYHGRVCVGILOSEAEILDEIVCPQOCSTEDAMTVLTPLT 2664
 Db 3826 --CERLDEDAFEKGDMP-----IPPEVTVEEYI-----DEHGHTVVKKVT 3867
 QY 2665 EK-----DYEGLKRVLRSLQAHKMAFPLEVPDNDAPDYGVIX 2704
 Db 3868 RKIIRIVYSEGTKEKEELMVQGMPO-----EPVNIIEGDGSKVIX 3908
 RESULT 10
 ID MYS2_DICDI STANDARD: PRT: 2116 AA.
 AC P08799; 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MYOSIN II heavy chain, non muscle.
 GN MHCA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87092266; PubMed=3540939;
 RA Warrick H.M., de Lozanne A., Lelwand L.A., Spudich J.A.;
 RT "Conserved protein domains in a myosin heavy chain gene from
 Dictyostelium discoideum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
 RN [2]
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
 RC STRAIN=AX2;
 RX MEDLINE=90353583; PubMed=2387408;
 RA Luack-Vietmeier D., Schleicher M., Grabatin B., Wippler J.,
 RA Gerisch G.;
 RT "Replacement of threonine residues by serine and alanine in a
 phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
 RL FEBS Lett. 269:239-243(1990).
 RN [3]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=88112226; PubMed=2828113;
 RA Wagle G., Noegel A., Scheel J., Gerisch G.;
 RT "Phosphorylation of threonine residues on cloned fragments of the
 Dictyostelium myosin heavy chain.";
 RL FEBS Lett. 227:71-75(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345066; PubMed=7619795;
 RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
 RA Rayment I.;

RT "X-ray structures of the myosin motor domain of Dictyostelium
 RT biochemiistry 34:8960-8972(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345067; PubMed=7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 RT truncated head of Dictyostelium discoidium myosin to 2.7-A
 RT resolution.";
 RL Biochemiistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=96206189; PubMed=8611530;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
 RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
 RL Biochemiistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RX MEDLINE=97452580; PubMed=9305951;
 RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMPNP complexes
 RT of the Dictyostelium discoidium myosin motor domain.";
 RL Biochemiistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=98070605; PubMed=9405148;
 RA Bauer C.B., Kuhlman P.A., Bageshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 RT of Mg.2(3+)-O-(N-methylanthraniloyl) nucleotides bound to the
 RT Dictyostelium discoidium myosin motor domain.";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS..
 CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDPA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14628; AAA33227.1; -
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MAA; 03-DEC-97.
 DR PDB: 1MMD; 17-AUG-96.
 DR PDB: 1MMD; 03-DEC-97.
 DR PDB: 1MMN; 03-DEC-97.
 DR PDB: 1MND; 17-AUG-96.
 DR PDB: 1MNE; 17-AUG-96.

DR PDB; 1VOM; 23-DEC-96.
 DR PDB; 1LVK; 28-JAN-98.
 DR DictyDb; DD01008; mhca.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAD.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00342; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
 FT DOMAIN 762 791 IQ.
 FT NP_BIND 179 2116 COILED COIL (POTENTIAL).
 FT DOMAIN 638 660 ATP.
 FT DOMAIN 738 752 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
 FT MOD_RES 678 678 ALKYLATION (SH-1).
 FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770B1EE56A1 CRC64;

Query Match 2.6%; Score 374.5; DB 1; Length 2116;
 Best Local Similarity 19.6%; Pred. No. 2.3e-06;
 Matches 347; Conservative 275; Mismatches 626; Indels 523; Gaps 78;

QY 31 DDDSDVPEEMEDDDDDASYCTESSF--RSHSTVSTPGRRKPRVHRPRSPILE----- 81
 DB 530 DEGSVFPN-----ATDWTLLTKLHSHS-----KKNAKYEPRRSKTEFGVTHYA 574
 QY 82 -----EKDIPPLE-----FKSSEDLAMPNEHIMANVIAIYELRNFGVLRIS 124
 DB 575 GOVYETIOWMLEKNKDPLOODELELCKDSSDNVYTKLFNDPNINASKAKGANFIV----- 630
 QY 125 PRRFEFCALVSEQCTLMAEMHVLLKAVLRBEDTSNTTFG-----PADLKMS 174
 DB 631 -----AAQYKEQLASLMAFL-----ETTPHVRVRCILPNKKDPAKLEK 670
 QY 175 VNSTYFIDGWTPEVLR-----VYCESDKYEYHHLVLP--OEAD----- 212
 DB 671 V-----VLDQLKNGVLEGIKITRKGFPRIIYADFVKYIYLLAPNVPADADSDQAKDA 725
 QY 213 -----YPGYVEKKIKVLOFLVTNNAREELMSEGV--IQYDHCRCVCHK 259
 DB 726 VLKHLNIDEGQYRFG-----ITKIFFRAGQLARIEARQRISETIKALQOATRGMIARK 780
 QY 260 LCDLLCETCSAVYHLECYKPLLEVPEDDEMOCEYCAHKVGVYDCAVLEQKN-KPYIR 318
 DB 781 -----VY-----KQAREHTVAARI-----IQDNLRAVI- 803
 QY 319 HEPIGYDSRRRYW-----FLNRRLIEEDTENEMEKKIWYSFTFYVLAELIDCDLK 370
 DB 804 -----DEKSWPMKFLKSKARPLKRNPERKEIKERE-----ILEKSNLTD--STQK 851
 QY 371 DYWEAELCKILEMREELHRRHMDITEDLTNKAAGSNKSFLLAANEELIESIPAKKGIDIN 430
 DB 852 DKLEKSL-KDTESNVLDIQRLK-AEKETLKAMYDSKDALKEQKRELEIRVDEMESEIDE 909
 QY 431 VK-SPEETEKDNTENDSKDAEKREFEED-----QSLKDSDD-----KTPDDDPDQG 479
 DB 910 KKLALENIQNOKRSVEEVYRDLEELQEBQKLRNTLEKTKKYEELMEKRVNDGSGDT 969
 QY 480 KSEVGDFFKSE-KSNGLSESS-----PGAGKGASGSTRITR----- 514
 DB 970 ISRLEKIDELQKEVEELTESPSESCKDGVLEKTRVALLQSELDDLIYVRLDSETKKSEL 1029

QY 515 LRNPDSKLSQKSOYAAAAAHEANKLFKEGKEVLYVNGGEISRLSTK--KEYIMKGNIN 572
 DB 1030 LRQKKLEELKOVQZALAAETAAKLAQERANKL--OGETELMEKNTSEVTASNV- 1085
 QY 573 NFKLQEGKRYRYHNOYSTNSFALNKHQHREDHDKRRHLAHFCITTPAGEFKWNG---- 628
 DB 1086 -----EKSKTLESQLVAVNNELDEBEKKNRDLAEKKAKALDAMLEEMDQLESTGEGKK 1139
 QY 629 SVHGSKVL-----ITSTRLLITTOLENNIPESFLHPWMAHRAVOMCKSP 678
 DB 1140 SLYDLKQESDMELRMOISELOSTIAKLEK-----IKST----- 1175
 QY 679 REFALALILECAVPRVYMLPIWREPLGHTRLHRMTSIREBEKYEKKKQKOE---DEE 735
 DB 1176 -----LEGV-----ARQGELEAQLAKSNVEKOKKKVVEILDLEK 1211
 QY 736 TMOQA--TWVKYTFPVKHQVMKQGEYRVYTGNGSMWISTKTHYRVFVKLPQNTNV-- 791
 DB 1212 SAQLAEETAAKQALDKLKKEQLESEVQ-----TOLSEAN-----NKNVNSD 1254
 QY 792 -YRKSLEGKNNMDEMDSDKRCSRKIKIPEDSKEDVKGSDAKADQNE--M 847
 DB 1255 STNKHLETSPNNLKLELEEQAKQALEKKRGLSEELKHVNEQLEEEKKQESNEKRV 1314
 QY 848 DISK-ITERKDDYVELLDSDSDKPCKEEPMVEVDDMKTESHVNGQESSQVYVNVSEGF 906
 DB 1315 DEKEVESLKDQILEEVVASKKAVTEKN-----KKESLDELKIQYADVVSBDKS 1365
 QY 907 --HRTSYTK---KTKSKLDGLER-----RIKQFTLE-----EKORLEKILEGKI 949
 DB 1366 VQOLTKLQAKNELNRTAEAGOLDRAERSKKAEFLDEAVAKNLEBETAKVAKRAM 1425
 QY 950 KGI-----GKTSNKKLSSES-----PVTKKEGQCSQSMQOBSP 987
 DB 1426 KKAETDYRSTKSELDADKAVSEQYVQIKRLNEBELSRVSEEDERONS-AIAAKTAA 1484
 QY 988 NANNQPEDLDIGCSQSSSVLRMS-----PSSHTTNK--LYPKDRVLD 1030
 DB 1485 ESALSLKDEIDAANNAKAKAEKSKLELVYAELEESLEKSGVYVNEFLRKNAELDD 1544
 QY 1031 V-----SIRPERK--CPKO-----NSIENDIEKVSDLASR 1060
 DB 1545 LNARLDRETESRIKSDSEKKNTRKQFADLEAKVEBAQREVVYTIIRLKKLSDDIISTQ 1604
 QY 1061 GQEPKSKTKGNDFTIDSK-----LA-----SADDIGTLCKKKKPLIOF-- 1101
 DB 1605 LDTETKSRK-----TEKSKKLEOTLAEERRAEEGSSKADE-----ETRKQVQDEVD 1653
 QY 1102 ESDTIYSSSKSALHSS-----VPKSTNDR--EGNYPLSRADFE-----G 1138
 DB 1654 ELRAQULDEBRALALNSEKKIKSLVAEVDVEKQLEDELILADKLVKARALEVELEVRD 1713
 QY 1139 KLGCSSESNTLENSD--TVSIDDSEEDMIVONSNEISIEQPTREQDYVLEPLKCE 1196
 DB 1714 QLEEBEDRSRSELESKRRLLTEVEEDIKKKYAEVQNTKLEAKKKLLDDVDVTLK----- 1768
 QY 1197 LVSGSESTNCEDRLPVKGTENGKKRPSQOKKLEEPVKKSGDQIKIKNTYTKKNENRES 1256
 DB 1769 -----KLEDEKKKLINESERAKRLEBENEDFL-AKIDAVAKNRSRAEK 1811
 QY 1257 EKKQORTFOINGKDNKPKIYLKGECLKEIYESRVVSGNVEPKVNNINKIIPNDIKSL 1316
 DB 1812 DRK-----KYEKDLKDFYK--LNDEAATK-TQTEIGAKKLEDDIDELRSLKLEBOQAKAT 1863
 QY 1317 TVKSARIPFINGVIMEDFNERNSETKSHLLSSDA-EGNYRDSLETLPTKRESQSTQ 1375
 DB 1864 QADRS--KTLLEGIDNLRQIIEDEGKIKMRLEREKRALBELEBELRETYEAEADSXS-- 1919
 QY 1376 TTTPSASCPSESNVQVDEMIETSE--VKVYSSPITSSEENLSNDPIDENG-LPIN 1431
 DB 1920 -----EADQSKRLVELELEDARRNLQLEIDAKETIADAKSNLQREIVEAGRLLEE 1970
 QY 1432 KNEVNGESKRRKTVITEVTTWTSVATESKTVIIVEGDKQOTVVSSTENCAC-----STV 1486

DB 1971 STARTSDSRKRLAEIALTAQVDAEQAKNOQIKENKIELEKYEKKFGESKTK 2030
 OY 1487 TTTTTLVTKLSTPSTGSGVDIIISVKPOSKTV 1517
 DB 2031 TKEFLVEKLETDYKRAKKEADEQOQRLTV 2061

RESULT 11
 ANK3_HUMAN STANDARD; PRT; 4377 AA.
 AC Q12955;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Ankyrin 3 (ANK-3) (Ankyrin G).
 GN ANK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=95138209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.;
 RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized
 at the axonal initial segment and node of Ranvier.";
 RL J. Biol. Chem. 270:2352-2359(1995).
 CC -1- FUNCTION: Membrane-cytoskeleton linker.
 CC -1- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in brain and other tissues.
 CC -1- SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13616; AAA64834.1; -
 DR HSSP: P55273; 1B18.
 DR Genew: HGNC:494; ANK3.
 DR MIM: 600465;
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000488; Death.
 DR InterPro: IPR000906; ZUS.
 DR Pfam: PRO0023; ank; 24.
 DR Pfam: PRO0531; death; 1.
 DR Pfam: PRO0791; ZUS; 1.
 DR PRINTS: PRO1415; ANKYRIN.
 DR SMART: SM00248; ANK; 21.
 DR SMART: SM00218; ZUS; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 21.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 DR Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
 KW REPEAT 73 102 ANK 1.
 FT REPEAT 106 135 ANK 2.
 FT REPEAT 139 168 ANK 3.
 FT REPEAT 172 201 ANK 4.
 FT REPEAT 203 230 ANK 5.
 FT REPEAT 234 263 ANK 6.
 FT REPEAT 267 296 ANK 7.
 FT REPEAT 300 329 ANK 8.
 FT REPEAT 333 362 ANK 9.
 FT REPEAT 366 395 ANK 10.
 FT REPEAT 399 428 ANK 11.

FT REPEAT 432 461 ANK 12.
 FT REPEAT 465 494 ANK 13.
 FT REPEAT 498 527 ANK 14.
 FT REPEAT 531 560 ANK 15.
 FT REPEAT 564 593 ANK 16.
 FT REPEAT 597 626 ANK 17.
 FT REPEAT 630 659 ANK 18.
 FT REPEAT 663 692 ANK 19.
 FT REPEAT 696 725 ANK 20.
 FT REPEAT 729 758 ANK 21.
 FT REPEAT 762 791 ANK 22.
 FT REPEAT 795 825 ANK 23.
 FT DOMAIN 1519 1898 SER-RICH.
 FT DOMAIN 4090 4174 DEATH.
 SQ SEQUENCE 4377 AA; 480399 MW; F42379E55768B684 CRC64;

Query Match 2.6%; Score 374; DB 1; Length 4377;
 Best Local Similarity 17.2%; Pred. No. 5,3e-06;
 Matches 526; Conservative 420; Mismatches 1080; Indels 1038; Gaps 116;

OY 17 QDSEDDDEDEMEEDDDSDY-----PEEMEDDDDA----- 47
 DB 1441 KETESDODDEIEKTRQSPASLALKRKRYSLTEPGKITERGTGATNSLPPTYSKPPFST 1500
 OY 48 ----SYCT-----ESSFRSHSTYSYTPGRKRP-----RVHRPRPILKEKDIPPLE 89
 DB 1501 RPYQSWTTATITVPGAPKSGFTSLSSSSSNTPSASPLKSIWSVSTP-SPIKSTLGASTS 1559
 OY 90 FPKSSEDLAVPNEHINVAIYEVLRNEGTVLRSPREPDPFCAALYSQEOCTLMABMV 149
 DB 1560 SVKSISDVASPP-----IRSLRTMSSPIKTVSOSPYNIQVSSGTLARAPAVTEAPPLKG 1613
 OY 150 VLKAVLREEDNSTNTPG-----PADLKDSVNSTLVFIDGMTPVELRVYCES 197
 DB 1614 LASNSTPSSSTSEVTTAGSLERSSTITMPRPASPKNTNM----- 1653
 OY 198 DKEYHNVLPRQ-----FEADYPRYGVENKIKVLOFLVDOFLTNIARELMSEGVIOYDH 253
 DB 1654 ---YSSLPKRSITTSAPLIPSSPLKSVSPKSRVDVISAKITMASLSPPVKQMPGH 1710
 OY 254 CRVCHKIGDL-----LCETCSAVYHLECVKPPLEVEPDEMOCEVCVAKKPGVT 304
 DB 1711 AEVALVNGSISPLKYASSSTLINGCKATATLQ-----EKISATNVSSVVSAAAT 1760
 OY 305 DCVAEIOKN-----KYIRHEPIGYDRSRK-----YWFNLRLIIEEDMENENEK 350
 DB 1761 DTYEVKFSSTTAMPSPSLRKYVSAPSAFOSLTPPSASALYTLGSSISATTSVYSIT 1820
 OY 351 KIWYYSF-----KVOLAELIDCLDKDYWEALCKILEEMREETH--RMDITEDLTNKAR 403
 DB 1821 TYPVYSVVNVLPEPALKLPDSDNSPTKSAALASPILKTLETENHPQHFRTSSPV---- 1876
 OY 404 GSKNSFLA-----ANDEILIESIRAKKGIDNVKS-----PEETE 438
 DB 1877 -KSLFLAPSAALKLSTPSSLSQSQELIKDVAEMKEDIMRTALITQDVPPEKPFQPELPK 1935
 OY 439 KDKNFENDSKDAEKNPEEEDDS--LEXD-----SDDK--FPDDDEPGKSE 482
 DB 1936 EGRIDDEPKIVKVEDLVKVSSEILKDYCVDNKSGSPKSPKSGHSEDEDMIESSSE 1995
 OY 483 VQDFSEKSENGELSESPG-----AGKGAS-----GSTRIITRLRNP--DSKLSQLK-- 526
 DB 1996 --EIREARQQAASQSSLSLPERVOYAKAKASEKDYNTKYVIDVLTNDIGSSSLTNLKYKF 2053
 OY 527 -----SQOY--AAAHAENKLFKEGKEVLYVNSQGEISRLSTK--KEVIMKGI 571
 DB 2054 EDAAKDGEGGOKRVLPALALQEHKLMKMPASMRSTSEKELCKMADSEFGCTJLILES-- 2111
 OY 572 NNYFKLQGEKGRVYVYNNQVSTNSFALNKHQREDHDKRRLAKFCLPAGGKRWGSGV 631
 DB 2112 PPDFSOHDODKSPSDSGFETRSEKTPPSAPQSAETTGPKLPEHVPVITTEVRYVH 2171
 OY 632 GSKVLTISTRLTITTOLENNIPSSFLPHNMASHRAWMKAVQMSKRPREFALALILECA 691

QY 2404 KHAVIHLKOKKSMTPAREENORMIYCNOMKYLIDKIDKEKOAKRRKRESEVOK 2463
 DB 4041 -----TSRGGQPSVTTKSAR-----DKTEAPLKSSEKAGSER 4075
 QY 2464 RSKQ 2467
 DB 4076 RSSR 4079

RESULT 12

MLP1_YEAST
 ID MLP1_YEAST STANDARD: PRT: 1875 AA.
 AC 002455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLP1.
 GN MLP1 OR YKR05W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RL Mol. Genet. 237:359-369(1993).
 [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354(1993).
 CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 REPAIR.

CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L01992; AAA34783.1;
 DR EMBL: X73541; CAAS1948.1;
 DR EMBL: Z28320; CA82174.1;
 DR PIR: S38173; S38173.
 DR SGD: S0001803; MLP1.
 KW Coiled coil; DNA repair.
 FT DOMAIN 69 487
 FT DOMAIN 531 1678
 FT DOMAIN 1834 1866
 FT CONFLICT 301 301
 FT SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match 2.5%; Score 360; DB 1; Length 1875;
 Best Local Similarity 17.9%; Pred. No. 6.5e-06;
 Matches 316; Conservative 298; Mismatches 656; Indels 494; Gaps 72;

QY 7 EEDDGAETQSEDEDEDEDDSDYPREMEDDDDDASCYCFSSFRSHSTYSSTPG 66
 DB 314 KEELNSIRLNTAKYADDSKOTPENEDLLKELQITKELKLAOCCEKCR 363
 QY 67 RRPVHRPSPILEKDIPLLEF-PKSESDLMV-----PNEHIMNIAIYEVL 115

DB 364 -----LSSITDDEADENENISAKSSSDFFELKKOLIKERRTEKHONOTETPIV-- 412
 QY 116 NGETVLRLSPRFREDCALVLSOEOCTLMAEMHVLLKVLREEDPSNTTFFGADLKDSV 175
 DB 413 -----ELEHKVPITNSFKERTDLENE--LNNALLLEHTSNEKNKAVKELNKK 459
 QY 176 NSTLYFDGMTPEVRLVYCESD-----KEYHNVLPYQEAEDYPYGPV-ENKI 222
 DB 460 NQKL-----VGCENDLQTLTKQRDLRQIOYLITLTVSVSDSKGFLKKEI 506
 QY 223 KYQLFLV--DQFLTNIAEELMSEGVIOYDHCRVCHLGLDL----- 264
 DB 507 OFIONIMOEDDSTPTESDSQKVTERLVEFKNIIOLEKNALIKVVRMLADLKSEK 566
 QY 265 -----CCENCSAVNHECVKPPLE-----EVP-EDPMQCEVCV 296
 DB 567 SKOSLOKISSETVNEAKEAITLTKSEKMDLESRIEQLKLEBELKTSVPNEASYSNVI 626
 QY 297 AHKVPQVDCVAEIQ-----KNKPPIRHE-----PIGYDRSR 329
 DB 627 KQLETKRDLSEQVQDLQTRISQITRESTENMSLKEIQLDLYDSKSDISIKKGKSSR 686
 QY 330 -----KYFELNRRLLIEEDTENENEKKIWIY-----STKVQDLIELDCLDK-DYW 373
 DB 687 ILAEERFKLLSTLIDLTK-AENDQLRRRFDYLONTILKODSKTHEYLTNEVSCSKSLIV 745
 QY 374 EELIKLEEMREIHRHMDITELTNKARGSKSP-----LAANEELIESIR-- 422
 DB 746 ETELLNKEEQKLRVHLEKNKQEL-NKLSPEKDSLRINVTOLQTLQKEREDLLETRRS 804
 QY 423 -AKKGD-----IDNVKSP-----EETEKDNE-----TENDS 448
 DB 805 CQKKIDELDALSELKETSQKNDHNTKOLEEDNNSNIEMQNKIEALKKDYESTVSDS 864
 QY 449 KDAEKREEFEDOSLEKDS-----DKTPDDP-----EQGK-----SEVG 484
 DB 865 KQPDIEKLQYKVSLEKEIEEDKIRLHTYVMDETINDSLREKLEKSKINLTDAYSQIK 924
 QY 485 DPKS-----EKSNGELSESPGAGKSGASTRIITRLRN-DSKLLKQGVAA 534
 DB 925 EYDOLYETTQSILQOTNSKIDES--FKDFTNOIKMLDEKTSLEKTSILKQ-- 975
 QY 535 HEANKLEKGEKEVLVNVNDSOGEISRLSTKKEVIMKGNIN--NYFKLGQGGYRVYHNOYS 591
 DB 976 -----MFLNNLELDQKKGKEKADFKKRISTLONNNEVEAVKSEYSKLSKINDLD 1030
 QY 592 TNSFALNKHQHRDHDKRRHLAHKFCFLTPAGEFKWNGSVHGSKVLTITRLITTOLENN 651
 DB 1031 QOITVANTQNNNEOELQKHADVSKTISLELR-----QLHTYRG-QVKTLLNSRDQLENA 1084
 QY 652 IPSSFHPMASHRAMVIRAYOVCSKPREPALA-----LAILCAVAPV----- 695
 DB 1085 LKEN--EKSNSQKESLELDLDSNRIEDLSQNKLLIDQIITYAADKEVANNSTNGRG 1142
 QY 696 -----VLPPIWREFLGHTRLHMTSIEREK--EKVKKKKEQBEETMOGATWVYTF 747
 DB 1143 LNNILTLRRERDILD---TKVTVAERDAKMLRQISLMDVBLDQARFKIDSRAVE--- 1195
 QY 748 PYKHQWKKQGEERYVTGIGGSMWISKTHYRFPVRLPGTNNVNYKSLGRTKNNNDENK 807
 DB 1196 KENHSSITIOOHDI-----MEKLNQMLNLR-----ESNTITLNELENNNKKKEIQ 1241
 QY 808 DESDKRCKSGSPKKIKIIPDSEKDEYKGSAAKAGADONEMDISKIT-----EKKDOVKEL 863
 DB 1242 SETDKLQNVAP-----ISESLTALKYSQEK--EQELKLKEEVRHMKKRSQDILE- 1291
 QY 864 LQSDSDPKCKEEMVEDDMKTESHVNCQESSQVDVNVSEGFHLTSTYKKTYSKSLDG 923
 DB 1292 -----KHQDLSSDYVEKLESEL--ENLKEELENE-----RQGALEAEKFRKLRR 1334
 QY 924 LERRIKQTLLEKQRLKLEKLGIGIKGIGTSTNSKNLSESPVITYAKKQCSQSDMKRO 983

Dh 1335 QAOERLTKSLSDSLTEOV-----NSLRDAKNVLENSL-----SEANARI 1375
Qy 984 EGSFNANNDDPEDLIQCSQSDSVLRMSDPSSHNTNKLYPEKRDVLDDVSTIRSEPKCPKQ 1043
Dh 1376 EELONAK-----VAQGNQNLLEAIRKIQEDAEKASRELQAK-----LEESTSY 1418
Qy 1044 NSTENDIEEKVSDLASGQEPSTKSKTGNDFFIDDSKLASADIGTILCKNKKPLQOES 1103
Dh 1419 ESTINGINEBITTL-----KEELIKORIOOOLQATISANE-----QNDL 1457
Qy 1104 DTIVSSSKSLAHSSVPKSTNDRATPLSRAMDEKICGCSSESNSTLENSSDTVIODSS 1163
Dh 1458 SNIVESAKKPFEDDKIKFTEKTOEVNEKILQEOERL--NOPSINIMEIKKKWESHEQ 1515
Qy 1164 EEDMIVONSNEISSEOPR--TRBQDVEVLEPLKCELVGSESTGNCEDRLPVKGTGANG-- 1219
Dh 1516 EVSQIKREAEALKKRIPLTEKINKIIRKKKEEL- EKFEKVEER--IKMEQSGEI 1572
Qy 1220 -----KPSQOKKLEERPVNKCSDI-----KLNKTTDKKNNE 1254
Dh 1573 DVALRKQLEAKVOEKELNEYNKIQEELKDPVHSSHISDDEPKLRAEISRLLEEF 1632
Qy 1255 ESEKKGORTSTFOINGKDNKPKIYLKGECLK-----EISRSRVSSGWEKRVNNINK 1306
Dh 1633 NNELOAIKKKSF-----EGKQOAMMKTTLLERKLAKMESOLSTKQSAESPXSNNV-- 1686
Qy 1307 IIPENDIKSLTVK-----ESAIRPFINGDVIME-----DENRNSSEKSHLSSSD 1353
Dh 1687 ---QNPLLGPRKLEIENSNSPFNPLLSGEKLTAKNSKSSGCGFNPTSPKNNHLONDND 1743
Qy 1354 AEGNYRDSLETLPSTKESDSTOTTTPSASCPESNS--VNOVEDMEITSEVKVTSSP--- 1409
Dh 1744 K-----REST-----ANKTPPTHLPEPSFNIPASRGLISSSTLSTPTNDEILTSNNPAOK 1794
Qy 1410 -----ITSEESNLSDNFIDENGLPINKNENYNGSK-RKTYITVTTM-----TSY 1456
Dh 1795 DSSNRNVQSGEDTEKKE-----GEVYKKEALIEEOTKSKNRPIDEVGLKDNEDDTTEN 1849
Qy 1457 ATESKTVIKVEKGQOTVVSSTEN 1480
Dh 1850 INESKKIKTEDEBEKETDKVNDEN 1873
RESULT 13
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD: PRT: 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATTRX (X-Linked helicase II) (X-Linked
DE nuclear protein) (XNP) (Znf-Hx).
GN ATTRX OR RAD54L OR XR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
RP VARIANTS ATTR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATTRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATTR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lissi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Coliaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATTRX gene encoding
RT a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Stayton C.L., Dabovic B., Gullisano M., Geetz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Geetz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA Millaesau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATTR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATTR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RL Submitted (Dec-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timst S., Villard L., Khrestchatskiy M., Fontes M.,
RA Coliaux L.;
RT "Specific interaction between the XNP/ATTR-X gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Buckmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Bickle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATTRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [9]
RP DISEASE.
RX MEDLINE=20213147; PubMed=10751095;
RA Villard L., Fontes M., Ades L.C., Geetz J.;
RT "Identification of a mutation in the XNP/ATTR-X gene in a family
RT reported as Smith-Rittenman-Ryers syndrome.";
RL Am. J. Med. Genet. 91:83-85(2000).
RN [10]
RP VARIANTS ATTR-X SER-1713.
RX MEDLINE=97196774; PubMed=9043863;
RA Villard L., Lacombe D., Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATTR-X phenotype
RT without alpha-thalassemia.";
RL Eur. J. Hum. Genet. 4:316-320(1996).
RN [11]
RP VARIANT JM GLN-2131.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Geetz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [12]
RP VARIANTS ATTR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aflimos S., Asenbauer B.,

Db 527 DIFENLETAMEVOSSVDHOGDSS--GTEOEY-----ESSS 560
QY 595 FALNKHORHEDHDKRRHLAHFCLTPAGEFPMNOSVHSGVLITSLRTITOLENN--- 651
Db 561 VKLN-----ISSKDNRGKIKSKTT-----AKYTKLYKLAPVSLPNSPIK 601
QY 652 -----IP-----SFLPHNMASHRANIKAVOMCSKPREFA-----LALATLE 689
Db 602 GADCOEVPORKDGYKSCGLNPK-----LEKCGLOEJNSDNLHLEVENESLLEE 650
QY 690 CAVK---PYWMLPIWREFLHTRLHMTSTIERE---KEK-----YAKKEK----- 729
Db 651 SDRSRPRVTTPLRRP-----TETNPVTSNDECKNETVEKOKLSVPRVKKDKRNSSDS 706
QY 730 -----KOEERTMOQ-----ATWVKYTFP 748
Db 707 AIDNPKPNKLPSKQSETVQNSDDEMLAILKGVSMHSSSSDDIDINEIHNHKLTYD 766
QY 749 VKHQVWK-QGGEERYVYGYGMSW-----ISKTHYRVPKLPGMTNMY----- 792
Db 767 LKTQAGKDKGKRRKRSSTGSDFTKKGSAKSIISKRRQTOQSESSVYDELEKEIK 826
QY 793 -----RKSLBGTKNMDENNDESKRK--CSRSPKKIKIIPDSKDEKVGSDAA 839
Db 827 SMSKIGAAKTTKKRIPYTK-DFDSSDEKHSKGMQNGHKNLKTQOEGSSD-----DAE 880
QY 840 KGADQNMEDISKITTEKKDDYVELDSDSPCKEPEEMVEDDMKTESHVNCQESSOVY 899
Db 881 RKQERETFSAGEVTDKDT IMELRDR--LPKKQOASATDQVDLKSQE-QSFTSLEY 936
QY 900 VNNS-----GFHRTSKYKKTAKTSKIDGLE-----RIKOPFLEE 936
Db 937 RKVAETREKSKHLTKTCKKVQ---DGLSDIAEKFLKKQSDSETSEDDKQSKQKTEEK 992
QY 937 KQ-----RLEKIKLE-----GGIKGIGKSTNSKKLSSPVI 969
Db 993 KKRSDPKKVKIKMEQOYESSDQTEKLPEREETICHPKGIKQIKNGTTOGEKKSKIRDK 1052
QY 970 TKKECCQSDSMKQEOSPMANNDOPD-----LQ-----OGCSQSDS 1006
Db 1053 TSKMKDELSTYAKSTGKSGCSDSEDKSKNGAYGREKKRCKLCKSSRKRRQDSSDT 1112
QY 1007 SVLRMSDPHTTKKILPKRVLDDVSIIRPETKCPQNSIE-----NDIEKYSDDLAS 1059
Db 1113 EKYSMEKDCGNS---DKRLRIELRERRLSSKRNRKEIOGSSSSDAEESSDNNK 1167
QY 1060 RQGEPTKSKTKGNDFIDDSKLASADDITLICKNNKPIQESPTIYSSSSALHSSVP 1119
Db 1168 KKQR-TSSKKK-----AVLYKEK-----KNSLRTSTK 1194
QY 1120 KSTNBRDATPLBRANDPEKLGCDSESNSTLNNSDTVSIOD-----SSEEDMIWONS 1172
Db 1195 RFOADTTS---SSSSDIE-----DDQNSIGEGSDQKIKPVTENLYVLSHTGCOSSG 1246
QY 1173 NESISEOFRTRODYV-----VLEPIKCELVSGSTGNCERLRLVYKGTENG 1219
Db 1247 DDAISISVAVYVDDDDDDNDPENRIAKKMLLEIRKANLSSDD-SSDDE-PEEKKKRTG 1304
QY 1220 KRPQOQKLEE--RPYNNKCSO-----IKLKNY-----DKNNENRES 1256
Db 1305 KONEENPGDEEAKNOVNSBDSDESSEKPRYRHLRLRKLVLVSGEGSEKKTTPKEIK 1364
QY 1257 EKKQGTSTFOJNGKDNKFIYLBCEKLEISERSVAGNVEPKYNNINKIIPENDIKSL 1316
Db 1365 EYKGRRR--KYSSBDSBDSDFQESGVSEVSES--EDEQPRPRRSKAKALEENQSY 1419
QY 1317 TYKESAIRPFIINGDYIMEDPNERNSFTKSHLSSDAAGNRDLEFLPSTKES----- 1371
Db 1420 KKKKRRRLKAYVEDSSSE--NKSNEEEEEKKEEEEEEEEEEDENDSKSPGGR 1477
QY 1372 -----DSQTTPPSASCPESNSVQVEDMEIETSEVKV-----TSSPTS-- 1412
Db 1478 KIRKILMDOKLRTETQNALKEEERRRRRIAREREREREKIRKAVITEIDASPLKCIYTKL 1537

QY 1413 -----EESNSLNDPIDENGLPIKNENVNG-----ESKRKT-----VITEVTT 1451
Db 1538 VLDEETKEPLVQVHNRMIKLRPHQVDQPMMDCCESYKTKKSPGCIILAHCMG 1597
QY 1452 MSTVATES--KTVIKYEKGDQTV 1475
Db 1598 LGKTLQVVSFLHTVILADKIDFSTAL 1623
RESULT 14
ID TRX_DROVI STANDARD; PRT: 3828 AA.
AC 024742;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z50038; CAA90349.1; -.
DR HSSP: P20393; IABY.
DR FlyBase: FBgn0014844; DVit.trx.
DR InterPro: IPR003889; FYRIC.C.
DR InterPro: IPR003888; FYRIC.N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001628; Znf_C4steroid.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00628; PHD. 3.
DR Pfam: PF00856; SET. 1.
DR SMART: SM00542; FYRIC. 1.
DR SMART: SM00541; FYRN. 1.
DR SMART: SM00249; PHD. 4.
DR SMART: SM00508; PostSET. 1.
DR SMART: SM00184; RING. 2.
DR SMART: SM00317; SET. 1.
DR SMART: SM00389; Znf_C4. 1.
DR PROSITE: PS50280; SET. 1.
DR PROSITE: PS01359; ZF_PHD. 1; 3.
DR PROSITE: PS50016; ZF_PHD. 2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator.
FT ZN_FING 1251 1334 PHD-TYPE 1.
FT ZN_FING 1335 1380 PHD-TYPE 2.

FT ZN.FING 1408 1469 PHD-TYPE 3.
FT ZN.FING 1708 1767 PHD-TYPE 4 (ATYPICAL).
FT ZN.FING 1768 1818 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-GLN.
FT DOMAIN 243 251 POLY-ALA.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 538 546 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
SO SEQUENCE 3828 AA; 413721 MM; 32059CF30A3C504 CRC64;

Query Match 2.5%; Score 358.5; DB 1; Length 3828;
Best Local Similarity 19.0%; Pred. No. 1.6e-05;
Matches 683; Conservative 439; Mismatches 1267; Indels 1215; Gaps 178;

QY 5 EEEEDGDAAEETODSEDEDEDEDEDDDDSDYDEEM-----EDDDDDASYCT 51
DB 523 EDNDDGSA-----SSDAIEDDEDIDDDAENEENEAASEKSAETASVDEKEADROLYW 577
QY 52 ESSF--RSHSTYSSTPGRRKPRVHRPRSPILEEKDIPLEFPKSSBDLWPNPHI-----M 105
DB 578 DNHFLPKRSTSS-----RIKPKRLLLEGIGICRSRSPDANGKRRKPKRYFGLATL 630
QY 106 NVAIAYEVLNFGTVL--RLSPREFECCALYSOEQCTLMAEMHVLKAVLREEDTSN 163
DB 631 PAKCTPRRRRSATATASOKLGKETEFASFATAYNSS-----FVLKOPRLQFOTDKS 681
QY 164 TTEGPRADLDVNSTLYFDGMTVREVLAVYCSDEKHHVLP-----YQEAADYRYGPV 218
DB 682 RSTVSA--KPLPTAT-----VLPASSAITSAVNLSEGL 715
QY 219 ENKIKVLOFLVQFLTNTIAREELMSEGYIYQDHCRCV-----HKLGLLCC 266
DB 716 NN-----ANSAVAAASTCAVCSAPVNNKNDAPLAKRYG-VIAC 751
QY 267 ETC-----SAVYHLECYKPRLEVP-----DEMOCEVCVAHKVGVTVCAVLEIOKNR 315
DB 752 EYCRKENSMTKISKLSTPMHNPSTSTASOGQOLKC-----TD----- 790
QY 316 YIRHEIYIGDRSRKRWFLNRLIIEEDTENEEKIKWYSTVOLAEILDCLDKDYWEA 375
DB 791 -----GGNC-----ILSLKSQLNKFKL--YKERCKACWMLKCLATLOLPA 830
QY 376 -----ELCKIL-EMREIHRHDITEDL-----TNKAR 403
DB 831 GHRSLATLIPRSMREBVAAPKDDKCELLSPASLFTAPTASSAGTITKMKSSAETAY 890
QY 404 GSNKSFLLAANE-----EILSIRAKGIDIDNVKSPEETE----- 438
DB 891 NSIKSNPLAENNVTFGCTPLRLPAILEKPLFLKIGSDNKKAKESKALGISPVTSEAA 950
QY 439 -KDKNTEENDSDAEKNREFEDOSLEKSDDKTPDDDEOGKSEVGDKSEKSNELSE 497
DB 951 VAPGKTTTRAKODKEKARLEAEKPLSPNAKKTTEANTPEOTKDEOPASTTTTVSASSS 1010
QY 498 SPGAGKAGSGSTR-I-TTRLRN-----PDSKLSQ--LKSOVAAAHEAN-----KLFK 542
DB 1011 TSHHTSAAATNSSOLETTEAANSAVPDNLKRORIDLKGPRVHVCSASIVIGOLPATFG 1070
QY 543 EKEVYLAVNSG--EISLSTKKEVIMKGINNVEKLIQOEGRYVYHNQ-----YSTNS 594
DB 1071 DEEBELAAAEAPAPATTTTTSPEYIIKRP-----KSPQMOMIDENDNCASCILTP 1125
QY 595 PALNKHQR-----EDHKRRLHAHKFCILTPAGEFKNVNSV--HOSKVLITLTLTLTT 646
DB 1126 EATAEQAPVAKSVLEBRSSKSNTOETAKRTPATISGSSKGKVTTRNATVTVSASSLVAT 1185

QY 647 QLENNIP-SSEFLHPNMA-----SHRA-----NWKAV-----OMCKSPREFAL- 683
DB 1186 KQORLENEVSSSISSQAAATQSRRLAKEVNRKLALISIDFMENTDPAEVCQ--TGFGIL 1243
QY 684 -----ALAL-----ECAYKPVYMLIWRKF-LGHTRLH--RMTS-IERE 719
DB 1244 VETVQORALCFJCGSTGIDPLIFCACCEPYYHQYCVLEYNLKHSSFEDTLTSLLETS 1303
QY 720 EKEKVKKEKKEQEETEMOATW-----VKYTFEPK--HOVVKOGEE-YRVTGYGWS 770
DB 1304 NNACALSAATNTALNOLTORLNLCPRCYCYTCNNSSGSKVACQKQKRYHSTCLGTSK 1363
QY 771 WT-----SKTHVYRFVYKLP----- 785
DB 1364 RLIGADRPILVCNLCCKSCATTKVSKFVGNLPMGTACFLKRRKGNFCIQCYDNDNF 1423
QY 786 -----GNTNVYRKSLBGTKN-----NMENMDESCK--RKSRSPPKIKIETPDEKDE 832
DB 1424 DLKMECGDCGNMNVHSCBGLSDEOYNLSTLPESIEFTCKKCARCQDYSRNKADEMROA 1483
QY 833 V-----KSDAKGADONEMDISKITEK 855
DB 1484 VMEEFKSLY SVLKLSKSKOACALIKLSPKRWKCCSAGAQAKAHSQ-----GLQPK 1538
QY 856 KQDVKELIDSDPKCKEPEMEVD-----DMKTESHVNCQESSQYDVVNVSEGFHLRTS 911
DB 1539 ALQFTYNGIGDSQESQNDIYEFKQHSNRKPRTPVPC-----SCLOPLSQSPFSL-YD 1594
QY 912 YKKTKSSKLDGLERRIKOFTLEEKQRLKIKLEGIGIKGTSTNSSKNLSESPVTK 971
DB 1595 IKOKIASNAYVSLAE-----FNYSQVIOO-----SNCELDIAVK 1631
QY 972 AKEGCQSDMSROBOSPNANDQ-----PEDLIGCQSDSSVLLMSPDSHTNKLXK 1024
DB 1632 -----ELSSQFPFQWQETACTDALEEDMFESCGYEELKESPTTYAEHHTASQAPR 1683
QY 1025 -----DRVLVDVSIKSPETKCPKONSIENDIEEKVSDLASGQEPTRSKTK-----GNDF 1075
DB 1684 TGLLDLPIDVDVLDG--GC-----AVKTRLDTRVCLFCFRKSGSLGEGEARLLYCGHDCM 1736
QY 1076 IDDSKLASADDIGTLCKNNKPLIOESDPTIVSSSKSALHSSVPKS-----TNDPAT 1128
DB 1737 VH-----INCAMSAVEFEEID--GSLQNVASAVARGMIKCTYCGNRGAT 1780
QY 1129 -----PLSRAND--FECKIGCDSSENSTLE--NSDPTVSIQDSEEDMIV 1169
DB 1781 VGCNVKSCGEHNYPCARTIDCAFLTDKSMYCPAHARMALKANGSPSYTESMEV----- 1836
QY 1170 QNSNESISQFTRQEDVLEPLKCEL-----VSGESTGNCEDR-----LPVKGTEA 1217
DB 1837 --SRPYVLEERKRK--KLIVAPAKVOFHIGSAVAVQDSIVYRFSDFAPYIPINFLCS 1891
QY 1218 -----NGKPSQOKLEBRPV-----NKCSQDQIL--KNTDKNNEBRESEKKG----- 1260
DB 1892 RLWYSSKEP--WKIYEYVTRTTIQNSYSTLILDLGRNFTVHTNPNCISLVQLOAQA 1948
QY 1261 -----QRTSPOINKDKKPKIYLKGECLKEISESRVYSGNVEPVNNTINKIIEPENDK 1314
DB 1949 RMHSLARSOLDLDTDAE--FPNSYVPAD--ENTEE-----EPO-QVADLLPPE--TK 1994
QY 1315 SLTVKESAIRPFG-----DVIM-EDFENRN-----SSETKSHLLSSDAEGNYRSLJET 1364
DB 1995 D-AIFEDLPHELLDGLSMIDPIFYEDLGKTELFAHSEBOSKOTTTATSAAG--ASVILC 2051
QY 1365 LPSTKESDS--TQTPPSASCPESSNVQV-----EDMEIETSEV-KKVYSSP----- 1409
DB 2052 DEDTRNSNSLNKHLVSNCTASNPVDAMLCAARSSOKECEGDVLLKKTDPATPSPMPK 2111
QY 1410 -----ITSEESLNDLFDENGLPINKNENNVGSEKRRKTVITEVTMTSTYATESEKTYI 1464
DB 2112 LDGGSVAARKRRRLSKNIAE--GVILLSLNO--RSKK-----EMATVA 2149

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:21 ; Search time 45.9589 Seconds
(without alignments)
5817.155 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 1433
Sequence: 1 MYSEEEEDDGAETQDSE.....RLKGFKASRSHNNKLGSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3363	23.5	810	2 G01252	small GTP binding
2	871.5	6.1	1711	2 T21432	hypothetical prote
3	522	3.6	2447	2 T16870	hypothetical prote
4	494.5	3.5	5327	2 T13564	microtubule-associ
5	459	3.2	5170	2 T15348	hypothetical prote
6	427.5	3.0	3759	2 A35085	trithorax protein
7	419.5	2.9	2232	2 T34434	hypothetical prote
8	413.5	2.9	3507	2 T34513	hypothetical prote
9	412	2.9	2897	2 B48666	cell proliferation
10	404.5	2.8	452	2 T21435	hypothetical prote
11	404.5	2.8	3256	2 A48666	cell proliferation
12	404.5	2.8	3488	2 T34418	hypothetical prote
13	402	2.8	1630	2 A53577	hypothetical prote
14	401.5	2.8	1367	1 S48478	glucan 1,4-alpha-g
15	399.5	2.8	405	2 T21433	hypothetical prote
16	399.5	2.8	510	2 T21430	hypothetical prote
17	399.5	2.8	2441	2 D71623	erythrocyte membra
18	397	2.8	2346	2 T13829	Trp homolog - frui
19	395.5	2.8	2271	2 F90073	hypothetical prote
20	394	2.7	2938	2 T30249	cell proliferation
21	387.5	2.7	1229	2 T25697	hypothetical prote
22	382.5	2.7	1589	2 T13606	hypothetical prote
23	382	2.7	2453	2 S60254	nuclear receptor c
24	375.5	2.6	3924	2 S37431	ankyrin 2, neurom
25	374.5	2.6	2116	2 A26655	myosin heavy chain
26	374	2.6	4377	2 A55575	ankyrin 3, long sp
27	373.5	2.6	2541	2 T29340	hypothetical prote
28	372	2.6	5105	2 T32650	hypothetical prote
29	367.5	2.6	3259	1 A56539	giantin - human

30	367	2.6	2215	2 T16871	hypothetical prote
31	366.5	2.6	3225	2 T52300	giantin - human
32	364	2.5	6642	2 T29757	protein UNC-89 - C
33	360	2.5	1875	2 S38173	myosin-like protei
34	358.5	2.5	3828	2 T13857	trithorax protein
35	356.5	2.5	2187	2 T30826	nascent polypeptid
36	355	2.5	1459	2 T32271	hypothetical prote
37	353	2.5	2035	2 A40718	host cell factor C
38	352.5	2.5	1871	2 D96796	probable heat shoc
39	351.5	2.5	3570	2 T45025	myosin MUC5B, trach
40	350.5	2.4	3147	2 T18674	hypothetical prote
41	349	2.4	2481	2 D90011	FnBb protein limpo
42	347.5	2.4	2364	2 A56577	microtubule-associ
43	347	2.4	1939	2 T18372	repeat organellar
44	347	2.4	2422	2 T12687	ALR protein homolo
45	345.5	2.4	1302	1 JG6009	surface-located me

ALIGNMENTS

RESULT 1									
G01252									
small GTP binding protein SEC4 homolog - human									
C:Species: Homo sapiens (man)									
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 29-Jan-1999									
C:Accession: G01252									
R.Bowser, R.P.									
submitted to the EMBL Data Library, January 1994									
A:Reference number: G06441									
A:Accession: G01252									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 1-810 <ROW>									
A:Cross-references: EMBL:U05237; NID:g1276427; PID:g451848									
C:Genetics:									
A:Gene: FAC1									
Query Match									
Best local similarity 23.5%; Score 3363; DB 2; Length 810;									
Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2;									
QY	1	MYSEEEEDDGAETQDSEDEDEDEDDDDSDYPEEMEDDDDACTESSFNSHT	60						
DB	1	MYSEEEEDDGAETQDSEDEDEDEDDDDSDYPEEMEDDDDACTESSFNSHT	60						
QY	61	YSSTPGRRKRVRRSPRIIEKDIPLERPKSSSDLMVNEHIMVATYEVLNRGTV	120						
DB	61	YSSTPGRRKRVRRSPRIIEKDIPLERPKSSSDLMVNEHIMVATYEVLNRGTV	120						
QY	121	LRLSPRFEDFGAALVSOECSTLMAEKHVVLEKAVLEEDTSNTTGPADLKDSVNSTLY	180						
DB	121	LRLSPRFEDFGAALVSOECSTLMAEKHVVLEKAVLEEDTSNTTGPADLKDSVNSTLY	180						
QY	181	FIDGMTPREVLRYVCSDESKYENHVLRYQEAEDRYRGFVENKIKVLFVLDQFLTNIARE	240						
DB	181	FIDGMTPREVLRYVCSDESKYENHVLRYQEAEDRYRGFVENKIKVLFVLDQFLTNIARE	240						
QY	241	ELMSEGVTDYDHCRCVCHKAGDLICSTGSAVYNHLEVCAPRLEVRDEMOCEVCANHY	300						
DB	241	ELMSEGVTDYDHCRCVCHKAGDLICSTGSAVYNHLEVCAPRLEVRDEMOCEVCANHY	300						
QY	301	PGVTDCAVETQKKRKYRHRHPIGYDRSRKRYWFLNRLLEEDTENNEKKIYYSTKVQ	360						
DB	301	PGVTDCAVETQKKRKYRHRHPIGYDRSRKRYWFLNRLLEEDTENNEKKIYYSTKVQ	360						
QY	361	LAELIDLDIDKYWEALCKLLEEMREIHRHMDITTEDLTNKAGSNKSFLLAANEELLES	420						
DB	361	LAELIDLDIDKYWEALCKLLEEMREIHRHMDITTEDLTNKAGSNKSFLLAANEELLES	420						
QY	421	IRAKGIDIDNVKSPETEKKDNETENDSKDAEKNREEFQDSLEKSDDKTPDDPEQCK	480						
DB	421	IRAKGIDIDNVKSPETEKKDNETENDSKDAEKNREEFQDSLEKSDDKTPDDPEQCK	480						

QY 481 SE----- 482
 Db 481 SEPTVEGDKGNSVANLGDNTNATSEETSPSGRSPVGLSETPPSSNAEKKVASEL 540
 QY 483 ----- 482
 Db 541 PDVPEEPNKTCESSNTSATTTSLIOPLENSSSELSNSESASAKADDPENGERSHT 600
 QY 483 -----VDFKSEKNGELSESPGAGKSGSTRITTRLRNPDSKLSQKSOOVAAA 534
 Db 601 PVSIOEELVGDFTSEKSTGELSESPGAGKSGSTRITTRLRNPDSKLSQKSOOVAAA 660
 QY 535 HEANKLEKEKEVLVNSOGELISRLSTKKEVIMKGINNYPKLGQESKRYRVYHNOYSTNS 594
 Db 661 HEANKLEKEKEVLVNSOGELISRLSTKKEVIMKGINNYPKLGQESKRYRVYHNOYSTNS 720
 QY 595 FALKHOREHEDHRRHLAHFECLTPAGEFKNNGSVHSGKVLTLTLRTITOLENNIPS 654
 Db 721 FALKHOREHEDHRRHLAHFECLTPAGEFKNNGSVHSGKVLTLTLRTITOLENNIPS 776
 QY 655 SFLHPNM 661
 Db 777 TSLHPSF 783
 RESULT 2
 T21432
 hypothetical protein F26H11.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21432
 R:Barlow, K.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19421
 A:Accession: T21432
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-1711 <MIL>
 A:Cross-references: EMBL:Z81515; PIDN:CAB04197.1; GSPDB:GN00020; CESP:F26H11.2
 A:Experimental source: clone F26H11
 C:Genetics:
 A:Gene: CESP:F26H11.2
 A:Map position: 2
 A:Insertions: 56/3; 154/2; 373/2; 417/3; 505/2; 596/1; 703/2; 847/3; 922/3; 1034/2; 1078/2;
 Query Match 6.1%; Score 871.5; DB 2; Length 1711;
 Best Local Similarity 17.7%; Pred. No. 5.5e-25;
 Matches 420; Conservative 299; Mismatches 618; Indels 1041; Gaps 71;
 QY 4 EEEEEE---DGAEEYQDSEDE---EDEMEEDDDSDYPEBEMDD---DDASTCTE 52
 Db 169 KROEEDIIYDEDESEEEBESSDDEFMLNEDQVVOEEBEEELNLDIKIEKGLDEENKCC-- 226
 QY 53 SSFSSHTSYSTPGRRKPRVHRPSPILKE--KDIPLEPKSSSEDLVPMNEHIMYAI 110
 Db 227 -----PWLDDEPRLPKLELPESSODPIPTAATIMOVEI 261
 QY 111 YEVLNFGVLRISPFREFEDFCALVSQEQCTFMAEMHVVLRKAVNEEDTSVTTEGPAD 170
 Db 262 YEILRSYHRLRTPFEFEEDFCALISHNNSCMAEYHMAALLRNLCKSDDEQTHSVTE 321
 QY 171 LKSVSVSTLFTIDGMTPVELRVYCESDKYHNVLPYQEA-----EDYPRGVENKTI 222
 Db 322 TNSVSIMIMHMDTLTYAELLROYIEA-----YFPADSVRDALINVDNYPFGVDANI 374
 QY 223 K-----VLOELVDFLTNTIAREELMSEGVIOYDHCRCRCH--LGDLLC 265
 Db 375 QROFSFEPNKAHVFLVLLFMSYREFLYSEFKLVNNGKPFONDECRVCGKSGSRVVG 434
 QY 266 CETCSAVYHLECYKPLLEEVPEDEMOCVCAH--VPGVTDCAETOKNKPYRIHBEPIG 324
 Db 435 CTGCEAAPHYEC--SHLKPRPE--VLVCNICKKNSAVRGVLPPEAVNDREP--LRSQPIGR 489

QY 325 DRSRRKYFELNRLIIEEDYENENEKKIWWYSTKVQLAELIDCDKDYMEALCKILEEM 384
 Db 490 DRYGRYWEIYRRLVYSLDETE-----LYYSTYPOLXOLQKDRYIEKDLCDTTRLR 545
 QY 385 REEIHRRMDITEDLJNKRAGSNKSFLLAANEELIESIRAK--GDIDNVKSP-EETKD 440
 Db 546 IDEFLQMALIVEMTSERR-----EALLETWVKROLIGYDPAEATTPQIYLHRD 594
 QY 441 K-NTEENDSDAEKNREFEEDQSLKSDDKTP-----DDDPQGSSEVGDCKSEKSN 433
 Db 595 SMKRMASILRDCAGKGVQKQEEVKEEPEVQGVQVYEDSILPESMGIF----- 647
 QY 494 ELSSEPGAGKSGSTRITTRLRNPDSKLSQKSOVAAAHEANKLEKEKEVLVNSQ 553
 Db 648 -----DAKL-----INTWMSG----- 659
 QY 554 GEISRLSTKKEVLT-----MKGNIN-----NYFKLGQESK---YRVYHNOYSTNSPALNKHQ 601
 Db 660 -----ATOELEVEQFVDISDNEDPAPSANLIRMGDEGNDQTFMYIYVYRNEMSESEFLT 713
 QY 602 HREDHDKRHLAHFECLTPAGEEFK-----NGSVHSGKVLTLTLRTITOLENNIPSFL 657
 Db 714 RKAADKKKTYMAKCF--AQIDNEDWVYAKNRQFYGDASLCKFTMTLQOVTINIPIDLM 771
 QY 658 HPNMAASHRAMWIKAVOMCSKPREFALALALECAVKEPVMIPIWREFLGHTRLRHMTSIE 717
 Db 772 HRKWPFEAKGFDELVSADYKKLVLTCLLDCAVRKTIFPMQMMNGLTGOTRLERITVDQ 831
 QY 718 REEKEKKYKKEKK-----QEEBETMOQATWYKYTFPVYHNOYKQGEERYTG--YG 767
 Db 832 RENFMKQOQRLKIDADALTKDLDSFVRVNYMKPKPNY--ILROGEFYRNAAGSGSMG 890
 QY 768 GMSWISKTHYRVFVKPLPGNTNVNYRKSLEGTNNMDENDSDCKRCSRPKIKIETPD 827
 Db 891 GMAWVAKKIYEKMI-----QYSEPKPLPLATYVEITTESY 926
 QY 828 SERDEVKSGDAAGADQENNDISKITEKQDQDYKELDSDSKPCKEPEMVEDDMKTES 887
 Db 927 SNRK-----ARLELIVSKIT----- 942
 QY 888 HVMCOESSQYDVVNVSGFHLRTSYKKKTSSKLDGLERRIKQFTLEEKORLEKIKLEG 947
 Db 943 -----KKRQSG----- 949
 QY 948 GIKGIGKSTNSKNTLSESPVITKAKEGCOSDMSROEQSPNANNQPEDLIQCSQSDSS 1007
 Db 950 -----GKSKKPTFFELT-----NGCYSPPCGRS----- 971
 QY 1008 VLRMSDPSHTNKLPRKRVLDVYSIRSPETKCPKONSIENTIEEKVSDLASRGQEPYS 1067
 Db 972 -----NPNRKC-----YSP----- 980
 QY 1068 KTKGNDFIIDSKLASADIGTLICKNNKPLIOEESDITVSSKSLHSSVSPKSTNDRA 1127
 Db 981 -----MCRN----- 984
 QY 1128 TPLSRAMDEGKLGCDSESNSTLNSDYSIODSSEEDMIYVNSNESISEQFRTREDOY 1187
 Db 985 ----- 984
 QY 1188 EYLEPLKCELVSESGTNCEDRLPYKGTANGKPKSQOKKLEBRPVNKCSDQIKKNTTD 1247
 Db 985 -----GYLVSAAQAHDERKLEB----- 1001
 QY 1248 KKNNEKRESKKQGRSTFOINGKDKPKIYLKGECLKLISRSRVYSGVNEPKVNNINKI 1307
 Db 1002 ----- 1001
 QY 1308 IPENDIKSLTVKESAIRPFLNGDYIMEDPNERNSSETKSHLLSSDAEGNYRDSLETLP 1367
 Db 1002 ----- 1001

Qy	1368	TKESDSTQTTTPSPACSPESNSVQVDEMDLEIETSEVKKYTTSSPTISEESENSDNFDIBNG	1427
Db	1002	-----	1001
Qy	1428	LPINKNENNGESKRRKTVILEVTYTMSTVAATESKTVIVKEKGDQKVSSSTENCASKSVNT	1487
Db	1002	-----SGVJGEBEK-----	1009
Qy	1488	TTTTVTYKLSPTSGSVSDIISVKEOSKTVVTVTVTDSLTTTGGTLVMSMTVSKEYSTRD	1547
Db	1010	-----	1009
Qy	1548	KVKLMSKSPKKTRSGTALPSTYKKEFVTKSTKKSIFYLPPDMLKTLARKGGIREV--PYFN	1605
Db	1010	-----KMPJEDTQTFSTKRGKGSIFLQKILIRIMMGCGOQVYMGEFS	1054
Qy	1606	YNAPALDIWPPSPRPTEGTRVRYLQGVKSLAGVLSLKLRLMALSRLMDMAAKVPPGG	1665
Db	1055	AGIKSNLLIMPYPAPRPTLDLCCKMOTLWARSLSHVAVALQKIMSISIFNEFD--PDOT	1111
Qy	1666	GSTRTETSET-EITTYTEIKRBDVPGVIREYCIKRIICIPI-GVPTPKETPTPKOKG	1722
Db	1112	HPDRVVIDPSPHDERRRIRIKHEMPYQGYERYEMETIEIPLVDEPEDEESMLSRNG	1171
Qy	1723	-----LRSSALRPKRET-----PKOT-----GPIYETWVAEELE	1754
Db	1172	GSSEFSHRSSARKKRPQRIHEFLSEKFGNYPKSKAFKSLDNRRATRIREWVDGVTLK	1231
Qy	1755	LME-----IRAAFEVVEKEKAQV-----EOQAKRLBOOKPVYIARSTTSPTSST	1800
Db	1232	VFEIKDYWKMKIRLEAETTKARKLEATRKAKAKAKEDERRRIDDQOORSVAKIPV-PMHS	1289
Qy	1801	TSTISPAQKYNVAPISGSV-----TTGTKNVLTTKVSGPATVTFQON-----KNFHOTF	1849
Db	1290	--LIPERNNV--PYLGSQOQRRPNNGENBERGLEKYNSSSVSPQAHGYASTPPEGHQPO	1345
Qy	1850	ATWKKQCSNSGVVQVQOKVLGIIPBSSTGTSOOTPTSPFORPATYVIRPNTSGSGCTTNS	1909
Db	1346	PNIRQAGYNO-----LPRKPTTSPFNFS--RP-VATIPPTPOLRAAGADG-	1390
Qy	1910	SOVITGPOIRG---MTVIRPLEOSTLGIKAIIRPPVWOGAPQOYWTQIIRGQ----	1961
Db	1391	--VVRAMVMTPRKNSIYNTSTPYPO-----ALNRQOYLOLQOQOQPAVRRLTNGYHMD	1443
Qy	1962	-----PYSTVASAPNTVSTPGOKSLTSTAT-----	1986
Db	1444	GTRMGGGRNPSYQMHQRLPQNRBAALQRPGESESTERRKRTTEALIRPDNGDEQRPVIRYD	1503
Qy	1987	STNSIOSSASOPRRPOGOYKLTMAOULTQTOGCHGNOGLYIVYOGQOTTGQLOL----	2042
Db	1504	PTSNFDORAKOQHPQSRPYVSTPAOMIRKTOGPGYKH--NYILMKASDGTQOKWLKRGQ	1561
Qy	2043	IPCGYTVLPBGGOOLMAAPNGCTVORFLFETPLATATASTSTTTTVSTLAAGTEGORS	2102
Db	1562	FPPGTVI--TSGQRVYVYRQPLTAVOQRQLYT-----ATPGTRVYRIRIPANGABRQDDH	1613
Qy	2103	KL-----SPQMOYHQDK--TLRPAQ-----SSSVAPAKAOPQTAOPSARP	2140
Db	1614	QVMRRVQVQASCPRAEMYMDQGTTPPGQOQRYVLAGGNSGTPDNVMPKVSSHGBRGLT	1673
Qy	2141	QPTQPOSPAPQEVQYQTPQVQOT---QTVSSNHPPSAQ	2175
Db	1674	MMQVQOQOQHNHPAYDMKPDATGFAVSTTEQVDPDQO	1711

RESULT 3
T16870
hypothetical protein T13H2.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C:Accession: T16870
R:Mu, X.
submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T13H2.
A:Reference number: Z18593
A:Accession: T16870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2447 <MAX>
A:Cross-references: EMBL:U039653; NID:g1049397; PID:g1049400; PIDN:AAB52494.1; GSPDB:G
A:Experimental source: strain Bristol N2; clone T13H2
C:Genetics:
A:Gene: CESP:T13H2.3
A:Map position: X
A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1
C:Superfamily: RING finger homology
F:158-207/Domain: RING finger homology <RNR>

Query Match 3.6%; Score 522; DB 2; Length 2447;
Best Local Similarity 18.5%; Pred. No. 9.2e-12;
Matches 502; Conservative 413; Mismatches 1033; Indels 760; Gaps 114.

```

0Y      9 EDQAEFTQJQSEDDDEEDMEDDDDDDDQDPEMEDDDDDAASCTRESS-----54
Db      2 DDSPEPSTSKARAKAEAAEAEENTSDSSDSSVSASASEBSRPSKKVYITRVLPVR 61
0Y      55 -----FRSHSTYSTPPRRKRPVHRPSPLEEKDIPLEFPKSSBDLWVPHENIMVY 108
Db      62 PTKDGHVYVNLSEGNSEETSLYQRAKEGIPSTKGRPEIKLPTTSQOYYDLEVLN-- 119
0Y      109 ALEYLRNFGTVLRLSPREFDFCALVSQOBQSLMAEMHVLLKAVLREEDTSNTFFGR 168
Db      120 -----PARMEG-----RELFLNA-----132
0Y      169 ADLKQSVNSTLYFLDGMATWPEVLYKYESDQKEYNHVLRYQEAEDYPRGVPENKIKYLOFL 228
Db      133 ---YDAVKNKNVNLPG-----KSVCEAD-----LOKV 156
0Y      229 VQGFITTMIAEBELMSEVLOYDHN--CRVHKIGLDLCCETCSAUVNLYSCVPRLPEVR 286
Db      157 IGSF--SCVOCEOLLQGSTMTKCGRRFCDOCLIVAFMSGNTCSTCRONIGSKRELQDP 215
0Y      287 E--DEMOCSEVCAHNKPGVATDCVAEIOKN-----KPIYRHEBIGYDRSKRWFLNRR 337
Db      216 RPDQLYIYQVESRSLYVG--RMAENREHEKUYVYGRKGYYIG--GDMWKKRGIDPNSK 270
0Y      338 LIIEEDTENENEKKI--WYVSTKVOLAELLDCDKDYWAELCKLIEEMREIHRHMDITE 366
Db      271 LKAPRPLKSAGRKKTRMFHES-----DED---GSVRKYVESKGA-----PKE 310
0Y      397 DLTNKAQSNKNSFLAANEBLLESTRAKKGOID--NKSPFELEKKNKETENDSKAEK 433
Db      311 DDTNLTENDKEGTSYAAKEVYLE-----EGEMDPRPEIKSSDEQOTLDBEESMLDSDF 365
0Y      454 NNEEPEDSL-----EKSDDKTPDDDEQDQKSEVGFQKSKSNGELSES 498
Db      366 ETSDEDEDKSPQCSSTKTTNRSRDSSSDSDSKDNELOKKKKRMAKKKNVPKTDGSDSVN 425
0Y      499 PGAGGASG---STRITRLR-----NEDSKLS--OLKSOQYAAAHANANKLFKGEKELY 549
Db      426 ESFDSDASGEVYATNLIESKSKKPPCGRRKKKFAPELLIEGDIPTPESDLSLSDDERDNA 485
0Y      550 VNS-----QGETSRLSTKEVYIMGNINY-----FKIGQEGKRVYTHNOYSTN 593
Db      486 ADPYAFVQKEFNRPDRDGHPEKDKLTNPFMDIMNHQVDRKEFEKGGEIHTVISD--DSN 543
0Y      594 SPALKHQRHEDHDKRRILAHKFCITPTGGERKWNMGVSNGSVLLITSLRLTIQJLENNIP 653
Db      544 S-----EHSDEADEDESSI-----DSEHEKKI-----566
0Y      654 SSFLAPNNAASHANMIKAVOMCSKPRFEALALILECAVKKPVVLPJWRFL-----GHTR 709
Db      567 SKFL-----SHR-----QPLPNPTS-----VDDCOVITVVKKVDYKQASILSKPETS 609
0Y      710 IHRMTSIEBEKEVYKKKEKKOEETWQQAIVWVYTPPVYAKHOVKKQGEIYRVLTGGM 769

```

```

Db 610 PDSSKIE-EKPDV--SEEVSDDEMPHITADKGDFTFLNIMHEHDEN-----YGC- 660
Qy 770 SWISKTHYRVPKLPNGTNNVNRKSLGKTKNNMDENDESDKRSRPAKIKIEPD- 827
Db 661 -----YLFK-----PGDTGIS-RPKVQBARAGTNRLSNNVCPAVALKEGKKLITPPEDY 708
Qy 828 --SEKDEYKGDAAKAGADONEMDISKITEKKQDVKELSDSPCKEPEMEVDDMKT 885
Db 709 EISDEVTYLLSDSEETSPSMEQSETSSEAGPTIITKSGTERTQSSSPFSTSRDR 768
Qy 886 ESH-----VNCQESSQVNVNVSSEGFHLRT-----SYKKKKKSKLD 922
Db 769 KMHKRLDTRRKRLADSDLSDFVFSIDGNEIYATGKPIIKHKVYDSANRMPKSNLD 828
Qy 923 GLERRIKQFTLEKQRL-----EKIKLEGKIGKIGKSTNSKSLSESPVITKAKEG 975
Db 829 FTGRNAREIPEEISRLAEQVAHEEYKIHRRQVYLAWEAASKLNYVDTEEBE- 887
Qy 976 QSDSMQEOGSPNANNOPELLIGCSQSDSVLRMSDPSTHTT-----NKLYPKRDVLDDV 1031
Db 888 -----IEEETPE-----EEVVK--VASPTAPLATENPTTSTAPFEEGVAMKETPEEI 934
Qy 1032 SIRSEPCPKONGIENDI-----EEKVSDLASRGOEPTKSKTKGNDFFIDSKLAS 1083
Db 935 FF-DPDEPCSSAQAOQRELIIEYKQOITIEDLEONKRP-SKT-----VAKESRE 986
Qy 1084 ADDIGTLLCKNKPDLQESDITVSSSKSALHSSVPKSTNDRATPLSRAMDEFGKIGCD 1143
Db 987 A-----QEBRIKDESEEQKKDADNPTEVDEKSEASSSDSDKDFE----- 1030
Qy 1144 SESNTLENSDQYSIODSSEDMIVONSNEJSEORTR-----EODVLEPLKCEL 1197
Db 1031 --DETIDQOSKYVILKHEKTV---SDEEI-EDFTRGFEVATADAKMKRTIGEV 1082
Qy 1198 VSGESTNCEDRLPVKGTENGKPKSOQKLEERP-----VNCSPQIKLNTTDDKNN 1251
Db 1083 VSTFEFL-----KLVAQPAVTEDEVLAGVCVNRTODEFSTIKETGKRTN 1126
Qy 1252 ENRSEK--KGRTSTOINGK-----DNKRIYKGECLKEISRVVSGNEPKVN 1302
Db 1127 KNPDESMVNFRES-FAAKHRVPKLPNIEIRMYI--ERAHNVKFKYHV--DMEDLHM 1181
Qy 1303 NINKIIPENDIKSLTVKES--AIRPINDVIMEDFNENSSETKSHLLSSDAEBNKR 1359
Db 1182 KILIALOKOITATCANLSPVTVTPPEEHAQVOLLHNLOPSTLRL-----NNPQ 1234
Qy 1360 DSELETPESTEDSTQTTTPSASCPEESNVQVEDMEIETSEVKKVTSPI-----TSE 1413
Db 1235 FALTILHAQOQOAIQOQAOQKQOKELARQOARVEELAKRKRIQOEAALAKQKGE 1294
Qy 1414 EESNLMDFLIDENGLPINKENNVGSKRTVITE--VTMTSTVATESKTVI-KVEKGD 1470
Db 1295 QMSNVS-----GIPVSSDQNAOSSNAQOTGLIENQTTTNSDSLTPRTLADNSHLGE 1347
Qy 1471 KO-----TVVSTENCAKST-----VTITTTVTKLSTSTG----- 1502
Db 1348 SQOIPVIESIOSTSEALKSESNYKMPILTPASTVSSKSAAPATRRPSRPPSSYDRPS 1407
Qy 1503 -----GS-----VDIISVKE--OSKTVVTTTVD 1524
Db 1408 PSVIVIRRLGSDALLNRPNRCNIDKSRSRSPISRAVETTVIINDGQOETLLAGNTH 1467
Qy 1525 SLRTT--GGTLV--TSMTVSKEYSTRDKVKLMEFSRPKTVSGTALPSYRKFTKSTR 1578
Db 1468 TVETITLEETISIGDSTIRYDEGCTYOID--KTIDDNSNGNINVDDEQSNVAK---- 1522
Qy 1579 KSLFVLPNDLKLARKGGIREVPYNYNAKPAALDIWYPSRPPTGCTWRKRLQTVKSL 1638
Db 1523 -----LRENDLMR-----EMLRAN-----RYHPSTMLAM 1547
Qy 1639 AGVSL-----MLRLMASLRMDMAKVPFGGSGTRETSETEITTTTILIKRDVAPYGI 1693
Db 1548 GNLSINERHNKVOOVLASOELDLARHTSGA-----VSOTOVE-----VGEG- 1592

Qy 1694 REEXCIRKIIIC-----PIG-----VPETPKETP-----TPORKGLRS----- 1725
Db 1593 -----VCAGTSDAIGETDEDDVDEEPEFVDDLELAKIKLOROGLESSESDSD 1643
Qy 1726 -----SALR--PKREPTRKQNG-----PVILETWAEELBELMEIRAPAEVY 1765
Db 1644 EDMVYDNDVDSVLRARPHKRRKTKKNIIVPNIPKIRKRYV-----DKKI 1690
Qy 1766 EKEKAQVADQOAKKRLDQOKRYATISTPTSTSTSTSTPAQVM-----VAPISGVY 1820
Db 1691 EMERA-----KYRARIKSQKMASTIRIAY--POKRTQOFATPOQVPRGKHSAAARAT 1743
Qy 1821 TGTMYLVITKVGSPATVFOON--KNFQTFATVWKGQSGNSGVNOYQKVLGIIIPST- 1877
Db 1744 PKPKAKMSVNSAPTKTITVAPNVAAGHQ-----NQQLXSDMAA-----POSTP 1789
Qy 1878 -GTSQOTFTSFPORPATVTRPNTSGSGTSSQVITGPOIRPGMTVITRPLQOSTLGR 1936
Db 1790 IRTPPOGTGSAPQAOQTPQSHLALQGFVNGAMQ--QABQ--QOGM--YTAQLQAMQ 1843
Qy 1937 AILRTPVMO-----PGAPQVMTQILRGQPVSTAVSAPRTVSTGQKSLTATSTSN 1991
Db 1844 AVAQTQAQAAVAEAAYAOVAQOARAPPOOLVQROYVGH--PGQVVPMPAOMLN- 1901
Qy 1992 QSSASOPPRPOGQVKTMAQLTQLOGHGNQGLVIVIGQOQTGQLDILPQVTVLP 2051
Db 1902 QGNPQMAVNAQOQMDERKMEV-----NAVYHMQSRQ-----EP 1940
Qy 2052 GPGQOQLMAAMPNGTVQRELFTPLATATTAATTTTSTVSTAAGTEBQOSKLSPOMYN 2111
Db 1941 PTNQELEVOVF--AOAQADLSAAAOAAQAOQMTNRAQAEAVARQAMMKQEOAR 1998
Qy 2112 ODFTLPPAOSSVGPAAKQAPQTAQPSARPQOPQSPAPQOPQOPQOPQOPQOPQOPQ 2171
Db 1999 -----AOAAKEAARILKAEIPEAKAKYQAEAEARRKKEQEKVR----- 2037
Qy 2172 SEAQPTHAAOSSKPOVAQOSPOGSPVRYVQSPQTRIRPSTPOLSPOQSOVQTT 2231
Db 2038 -QAOAQTQAQAOQAOQAOQAHQAOQAOQAOQAOQAOQAOQAOQAOQAOQAOQAOQ 2074
Qy 2232 TSQPIPIQPHITSIQISQOQPOQOPOVQOSTQPLSSGQILNVS-----VSSPSRPLQ 2285
Db 2075 SGQPLMQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2134
Qy 2286 IOOPOP--GVIAVPQLOQOQOVLQLOQOQVQAOI-QAOQSGVPOQIKLPLQIOQSSA 2341
Db 2135 VQOGMPAGIOMFTGMQOQLOGLMPGMQOLPQOAGQSOQTSOAOQOQLFMOQLQOQOL 2194
Qy 2342 VQTHQIONVTVQOASVQEOQLQVQOQKQOQKQOQOQOQOQOQOQOQOQOQOQOQOQ 2396
Db 2195 MQ-HQOQOQLOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2253
Qy 2397 IQOQVYMK 2404
Db 2254 VQOQNLMR 2261

RESULT 4
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
M:Alternate names: hypothetical protein EG:49P4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>

```


QY	1776	QAKRLLEOKPVLVITSTSTPTSTISPAQVVMVAIPSSGVTTGTKVLTKYGS--P	1835
Db	3500	ESRREVAEKSLPEKEKSRPTVAEYVKDAEKSKEESRRESE--AEKSSLSAKASRP	3557
QY	1835	AVVTQDNKNFHPQTEPATVWKQGSNSGVVQVQVVLGLIIPSSGTQSGFTTSPQETAV	1894
Db	3558	ASV-----AESVDEAEKSEESRRESEVAEKSPPLAKES-----RPASVAE	3598
QY	1895	TTRPNTSGSGGTTNSQVITGPQ-----RPGRVTRVPLQOSTGLKAIITRPVWQPG	1944
Db	3600	SVKDEAEKSKEYSRRESVAEKSPLPEKASRP--TSVAEYVKDAEDSKESR-----ESG	3654
QY	1949	APQOVWTOIIRGQPVSTVAVSAPNTVSTPQCKSLTSTSTSNIGSSASQPPRQGGQVLT	2008
Db	3655	AKKSPSLASMEASRPISVAEYVKDETEKSKEESRRESETEKSPPLPEKASRPISVAEYK--	3711
QY	2009	TMAQLTQLTGHGNOGLTVVYIGGGQTTGQLDLIPGVTVLPFGQQLMQAMNGTVQ	2066
Db	3714	-----DEAEKSEESRRESVAE	3730
QY	2069	REFLPPLAT--TATPASTTTTSTTAAGTGEORQSKSLPQMOWNQDKTLPPAOSSVGA	2122
Db	3731	K---SPLSKESRSPASVAEIKDAAGTQKESRRESMPES--GKAESTIKGQSSLSAKE	3788
QY	2128	KAQPTAPASARPOTQPSA--QPEVOTPEVOTQTSSVHPSEAPTHAASRQ	2189
Db	3786	TSPRPSVASEYKDFE--KPEASLADKQVASRE-----SVANSADKESPLH--SRPE	3833
QY	2186	VAAQSQPOSNNVQGSVPKVGQSPSGTRIR--PSTPSOLS-----PQGSQVQVTTTSQPIPI	2236
Db	3837	SVADKSPASKAEKASRLSVATASPPLEEGRSLADSLPLNLTGEAKGLPTLSPIDV	3896
QY	2239	QHTSLQIPSGGQPG-----SQPGV--QSSQTQLSSGQTLNQVY-----	2278
Db	3897	AEGDPLEVAKASSPRAVLSKPAESQPDGTGHTASTPVDASVLEIEVWBOHTTSGVG	3956
QY	2277	---SSSPRQQLQIQPPQVLAIVPQLQOQVYLSQISQVAAIQAGQSGVPG--QIKTLQ	2332
Db	3957	ATGATAETDLDLDETSEYVT--KQSETTLEETLSKVESKVEVLESVQVGEKQYT	4013
QY	2333	PI-----QIQSSAVQTHQIQNVVTVQAAVSQEDQLQ--RVQQLRDQQKKKQOQ	2379
Db	4014	SVKQAEVTTVDSLRLQTKKSESEQLTEIKSVLDPININVTNLSTAVETIEKKVQDVTEKV	4073
QY	2380	IEIKKEH-----TIQASNOSEIIQKQ-----VYMKINA	2407
Db	4074	IEKATEHYSEHVTTGSESTSEQKSSLDLGFSELRETHITTVGSPREFTVITCERDP	4133
QY	2408	VIEHLKQ-----KKSMPAE-----REENOMITVCN--QVMKY---TLD	2441
Db	4134	VLHDIKEDEEHRRSPSPDVDAKAIIPQGMRLPSREEEYAKVIADVAKVILSKDQLTD	4193
QY	2442	KI--DKEEKQAAKKKKRESEYVQKKRKNQATKLSALLPKHKEQILRAELIKKRALLDK--D	2497
Db	4194	IIPDEQDLQEEKLKSTADTEEEESDKSTPDEKLEI-----SVKVEIEEKSPPQKSGP	4248
QY	2498	LOIEVQELKEDLLIKKKEKDLQALQATVAACAPRVTVLVLRAPARPPSPRRPGVQHT	2557
Db	4249	ISTEKKDITQESKAQLRQGLASSRRESVA-----SQPSVP-----	4286
QY	2558	GLLSTPLPVASQKKRKEEKSSSKKKKKMISTSTKTKKDKT	2602
Db	4287	-----SPQSASASHHKLEVELSESHKAKKESRSPASVASYSEKDMK	4327
RESULT 5			
T15348			
hypothetical protein B0350.1 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999			
C:Accession: T15348			
r:Galtung, S.			

[illegible]

```

Db 2388 PSEDESSRVETTTTTVTREHFEPEDDHS PVVQ-----TOEYASSES 2430
OY 724 VKKKEKROEE-ETMOQATVWKTPYVKHQVWKOKGEEYRVTCYCGMSLSTKHVY RVP 782
Db 2431 PVPSEKSEVERIETTTTTVTREHFEDEHIIQCOGES----- 2468
OY 783 KLPQNTNVRKRSLEGTKNMNMENMDSKRCSPKPKIKIEPDEKDEKVGSDAKGA 842
Db 2469 -----DSQIPSEITSSENM-----RETSSSPVQ-----SNRDE----- 2497
OY 843 DONEMDISITEKKODVVELDSD-----SDKPCKEPMEV---DDDMTSHVNOEBS 895
Db 2498 ---EVLPAIARYKOPTEKGRVDSHDAPASAPASAPISQAYKQESOEKHSLEOQOR 2554
OY 896 QVDVVA--VSEGFHLRTS--YKKTKSKLIDGLERIKOFTLEKORLEKIKLEG---GI 949
Db 2555 SVSHEDSPAQOYFHNDSEHSDSPSPDRAPLITEOHOPESGESDG-----EGFGSKY 2608
OY 950 KQIGKTSTNSKNLSSEPVITRAKEGCG--DSMRQOSPNANDOPEDLIQOCSOSDS 1007
Db 2609 LCFAKKAGVAGVAVAPV-ALAAVAKAAYDALKKDDDEDENDPEQKLP--KSPERQ 2665
OY 1008 VLRMSPSTTKLYPKVDVLSIR---SPETKCPONSIENDIEEKVSDLASRQOE 1063
Db 2666 VL-----VMPVEPSEISEIELEYTSPSEKSESQCYTE----- 2701
OY 1064 PTKSKTGNDFFID-----SKLASADDIGTLCK-----NKKPLIOESDITVSSSKSA 1113
Db 2702 -IVRTTIVREIIDDQOSTVRSRSPSEHDISEOYAPESPVEDPPIVETKTYVIRQYHDE 2760
OY 1114 LHSVVKSTNDRAPPLSRAMFEGKLGCDSESN--STLENSSDTVSIDSSEDMIVQ-- 1170
Db 2761 PPOEIEQOTIPEBVYLVREV--YESPEGDEPEQHIETKTTTITKEVHVAVEDVQISR 2818
OY 1171 -NSNEISF-QERTRO-Q-DVEVLEPLKCELVSESTGNCEDRLPVKGTFRANGKPKSQOK 1227
Db 2819 VHSETSVSSEKQPADQLEPTESTTATVREYEEPEEVPSPGSEADE--SHAQ 2876
OY 1228 LBERPVNKCSDQIKKNTDKNNRESEKKGOPTSTPOINKDKPKITLYKGECLKET 1287
Db 2877 YMETT-----TTTIVTREYEVSEDEHQSOV-----QRDSPA--PSEDSVKHY 2920
OY 1288 SE-----SRVSGNVEKVNINKIIPEND-----IKSLTVKESAIRPFIINGVIMEDFNE 1338
Db 2921 IKETTTTIVTERYBE--DSHSPVSEDDVHGFKVTTTTT-----VTHHFEP 2969
OY 1339 RUSSEKSHLLSSDAEGYRDSLETLPSTKESDT---QTTTPASCPSPNSVNOVEDM 1395
Db 2970 EDPSPDEHVESEERYASGS-----PVPSEDSREIETTTTTVTREHFELEDDEH 3021
OY 1396 EIESVVKVTSSPITSEE-----ESNLSN-----DFIDENGLP-----I 1430
Db 3022 VVESQEV-SASSPVPSEKSEVERIETTTTTVTREHFEHEDDIPITIVETSHDDPAASSV 3080
OY 1431 NKENVNGESKRRKVTIVETWTSTVATE-----SKTVIKVEGDKQIVVSTENC 1481
Db 3081 PSEEDVHGQ-----IQTTTITTVTRHVVPDEIDSGRMDLEKYSSESVPSEDS 3133
OY 1482 AASTYTTTTTITK-----LSTPSTGGSVDIISYKEDSKIVTTTTVDSLT 1527
Db 3134 SVRIETTTTTVSRHEFEDEDDSHVSESQESASGS-PVPSEKSEVERIETTTTT----- 3188
OY 1528 TTGGTLVTSMVSKSESTPDKVKLMKFSRPKJTRSGTALPSYKVEYTKSKKSIFLAPND 1587
Db 3189 -----TYTRHEFTDEYI-----PSESRT-----SHDDGITDOHVPQSQSPVSE 3228
OY 1588 DUKLARKGIREVPY-----FYNAPALDIMPSPRPFTGI----- 1626
Db 3229 EDDHTHEQIINDPIDEQIVESHKESST-----PSEQOSTHYIETVMSTPINSERY 3283
OY 1627 -----TMY-----RLQTVKSL----- 1638

```

```

Db 3284 DPEVEKDVESADEIDISSTAQYKSESPVOTEKSIILLAKOQOESGDESGFGSKVLG 3343
OY 1639 ---AGV-----SLMLRLMASLRMD-----DMAKVPPOGG----- 1666
Db 3344 FAKKAGVAGVAAVAAVALLAAVAKAAYDALKKDDDEDEDEKBPILIGCHKQODPI SD 3403
OY 1667 ---STRETSIEITTE----- 1681
Db 3404 EPSAVSETOPEPATYEPREEDKVIITDSADSVQDEPKIVPVDSTPEHNSNDREEFE 3463
OY 1682 -IHKRDVGRYGIREFYCIKTIICPIGVETPKE-----TPTRORKGIRSSALRPKR-- 1732
Db 3464 STVKSE--GPY-----IVESTDYAQTSAEPRISSPVHSDAGSEFRFPESVTG 3511
OY 1733 -----PETPKOTGPVLIETVWAELEL-----WEIRAFAEYERKEKAQAVQOAK 1778
Db 3512 EDEKNAIPEISEIDAYVIDS---EYERNNNDQORISSPAISDEDEDAVIDISEFY 3567
OY 1779 KRLQOK---PTV-----IATSTSP--TSSTSTI-----SPA----- 1807
Db 3568 RHSEOQNNEDPSSIVSGEYISSGHGSPREFEDSTTTVLTNVHNEPAALPEPEVEDELE 3627
OY 1808 -QKVMV-----APISGVTTGKMLVLTKY-----GSPATVTPQOKNFHOTATWY 1853
Db 3628 QERSIIESEYKTSPLPPTSVTVHEVPEALIKHYRTSPRTIVTVSEHLDRODEPY 3687
OY 1854 KOG-----OSNGVVOOVOKVLGIIPS-----STGT 1879
Db 3688 VESEETTRASPLGERPESPSSGPLREEDSHVISHETSSPVSEDSVKVYIEKTT 3747
OY 1880 SQOFTSPQPR-----TATVTIRPNTSGSGTTSNSQVITGQIIR 1920
Db 3748 TVTVERYEPEDSHSPVSEDDVHGFKVTTTTTIVTNEHFEDEHDSHVESERYAS 3807
OY 1921 GMTVITPLQOST--LGKALIRPVAVQCARQOVMTQIIRKQPVSTAASAPTVSGTB 1978
Db 3808 GSPV---PSEENSNRYETTTTTVTREHFEPEDDSHVESQESYA-----SGSPVPS 3858
OY 1979 OKSL-----TSATST-----SNIOSSASQPPRP--OGVKLMAQ 2013
Db 3859 EKSEVYIETTTTTVTREHFEHEDIPITVESHDDQAASVPSEEDVHGQIOL-ITTTT 3917
OY 2014 TOLTOGHGNOGLTVVIOGQOQTGOLQIIPGVTVLPGGQOLMAAMPNGTVOREFL-- 2071
Db 3918 TVTTRH-----VVPBEIDSGRMDLEKYSSESVPSEDSRVLETTTTTIFIR 3969
OY 2072 -TPP-----LATTTATAS-----TTTTVSTTAAGTGQORQSKLSPOM 2108
Db 3970 HFEPEDDSHVSGOERYASGSPVSEKSEVERIETTTTTVTREHFEHEDHA----- 4023
OY 2109 QVHODKTLPPAOSSVGPAPKACQOTLOPSARPQOTPOPSAPPEVOTQOPEVO--TOTTV 2166
Db 4024 QILD--TVSEKTSSEPEVQITLEADITBETHORSABEDHOP--ISTEKPVHGMETTS 4080
OY 2167 SSHVSEAPTH-----AOSKPOVAASQPOSNVQO--QSPV----- 2202
Db 4081 TSHTASHVELEHDDDESGEGFGSKVLGFAKKAGVAGVAVAAVALLAAVAKAAYDA 4140
OY 2203 -----RVQSPQO-----TRIRPSTQSLSPOQSO---VQIT 2231
Db 4141 LKKDDDEATYDLDRHGVSYSALQRPQOVESROEDSESGSPFDHGDSHLEVOQT 4200
OY 2232 -----TSOPIPIQPHSLQIPSOQPOQPOVQOSTQTLSSQOTLNOVSSPSRPOQ 2285
Db 4201 EAAFKLSSEDIABQ--TAEKLVEVAPLDRILISEQIRASIESLNR--NSPVEPKNS 4255
OY 2286 IQQPOPOVIAVPOLO-----QOVYLSQIOSQVAVQ 2316
Db 4256 VED-----LASEALQVNTVTSFEGAHESDDKHDGNDGEMKYVDRNGEVLEEFSTGLTD 4310
OY 2317 I--QAOQSGVPOQIKLOLPIQIOOSSAVOTHQIONVYVQAASVOBOLQRVQOOLRQOQK 2374
Db 4311 VIOEAGDATTHIMTOAEYSPRRSKFLKQESCOEISNEPEVDYSDLO--EKLTILAGE 4368

```



```

Db 1671 LEEDMESSCGGVEDLODAGVSASVNEHSTQAEBSNGVD-----IPL--- 1717
Oy 1046 IENDIEKVSDLASRGOEPTKSKTKGNDEFIDSKLASADIGTLICKNK----- 1095
Db 1718 -----EEVDDGSGGCIK-MRLDTRMCLFCRKSGEGLSGEAFLLYCHDCWHTCAMK 1770
Oy 1096 KPLIOESPTIYSSSKSALHSSVPKS-----TNDRAAT-----PLSRAM 1134
Db 1771 SAEVEEIDP-----GSLQNHVHSAVARGMKCTVCGNRGATVGCNVRSGEHHYPCARSI 1826
Oy 1135 D-----PEGLGSDSES-----NSTLENSDPVSIODSSEEMIVQNSMESISEQRTREQOV 1187
Db 1827 DCAFLTDKSMTCIPAHKNCNALKANSGPSVYTESNEV-----SRPYVEIDRRKK--- 1877
Oy 1188 EYLEPLKCELVSG-----ESTNCEDRLPVKGTGA-----NGKKPSQOKLEER 1231
Db 1878 KLIIEPARVQFHIGSLLEVROLGAIVRPDSQYEAIVNIFLCSLNYMSKEP---WKIIVEI 1934
Oy 1232 PV-----NKCSQOIKL---KNTDKKNNESEKKQ-----RTSTFOINGKD 1272
Db 1935 TVRTTIQNSSSTLALDVGRRNTYVDHTNPNSKEVQLGMAQIARMHNTSLARSEFLENGTD 1994
Oy 1273 NRPKIYLKGECLKEISESVSGNVEPKVNNIKIIP---ENDIK---SLTVKSAIRPEI 1327
Db 1995 WSGEPPNPNSCVPP-----DQNTEEEPQOQADLLPWLITSPKFLGLSTHGILLML 2047
Oy 1328 NCDVIMEDENERNSETKSHLSSDAEGNYRDLSTLP-----STKESDST 1374
Db 2048 LGVVV---RLNQGELKNAIF---EDLPHEILDGISMILFIYLDKTDLFALSEQSKDGI 2101
Oy 1375 QTTTTSASCPESNVQVEDMEIETSEVKKVTSSPTSEBSNLSNDFIDENGLPI-NKN 1433
Db 2102 QAMTS-----NQAOONOAGANSVS---ICDEDFRNSNTSLG-NGMPANPV 2146
Oy 1434 EVNGESKRKTYTEVTMTSTVA-----TESKTVIYVEKDKQTVVSTENCASKTV 1486
Db 2147 EDMMLSAAN---SSOVOMLKTAMPKLDGNSMATAIKRRKLK-----NLAGEVF 2195
Oy 1487 TTTTTLVTLSPSTGSDIISVKEQSGTVVTVTDTLTTTGTLVTSMT---VSEKEYS 1544
Db 2196 LTLSSQORRKKEATYAGV-----SRROSISTSVGEVATTSGSVRSKSFYSAKRRIF 2249
Oy 1545 TRD-----KVKLMKESPCKTRSGTALPSYRKFTVTKSTKSIIVLPNDLKLKARKG 1597
Db 2250 EKSEGESEAARIMQMDG-----VDDSTIEFRISGONISTAQFSGQ 2293
Oy 1598 IR-----EVRFYNNAKPALDIMPSPRPTFGITMRYRLQTYKSLAGVSLMLRLNASLR 1653
Db 2294 VQCDRCQCTYRNDAFOR---HLPSGSPT-----MSSNETESDVS----- 2330
Oy 1654 WDDMAKVPBGGSPTRETSSETITTEITELIKRDVGPYGRFEYCIKIIICPIGVPETPK 1713
Db 2331 -----GGMTNNATQISNESINELQKOLLANAGGLN---CLQ-----S 2365
Oy 1714 EPTTPQ-----RKGLSSALRPK-----PETPKQTPVLIETVAEE 1751
Db 2366 ANSPFQVQSLGSLGRLGLOLQLOLQPSLGDGFFLSQPDATQADT-----D 2414
Oy 1752 EELM---ETRAAERY-----EKEKAQAVEQOAKRLBQOKPTVATST 1793
Db 2415 DLQIYANSLSQSLAANGGFTLAOPTVTAPOQLIAVSTNPDGTOQFIQIPQOMATTT 2474
Oy 1794 TSPTSTSTISPAOKVWAPISGSVTVTGKMYLT---TVGSPATVTPPOONKNEHOTPAT 1851
Db 2475 PTAITVOTLQATITDKIML-----PLTAAGKPLKTVATATAAQAQAAVQKQHLASGHQVKPI 2529
Oy 1852 WVK-----OGQNSGVVVOQ---KVLG-----IIPSTGT-----SQQFTS 1886
Db 2530 QAKLDRPQHQHQOQOQVQOQVITVWGQNLQRLQFLFQSGSTQOAPQIILLPQAQPNITIS 2589
Oy 1887 F-----OPR-TAVVTRIPNTSGSGG-----T 1906

```

```

Db 2590 FYTGDSQGOQLOLYISIPITAGEKKPQOPQIPATPTFLTTPAGACATVLTQDASNLVLT 2649
Oy 1907 TSNS---QVITGPQIRPGMVTIRTPLOOSTL-----GKAIRTPVMVOGAPQOVM--- 1954
Db 2650 PNSSGLQMLTAQSLQAPQVIGLIRPQITQLGAGDGN-----QPGSNQDPLILG 2700
Oy 1955 -----TOIRGQPV-----STAVSAPNTVST-PC---QKSLTSA 1985
Db 2701 GNGGSSGLEFATTSPOVILIATQPMYGGLETTIVQNTVMSSQCFVSTAMPQMLSQNASFSA 2760
Oy 1986 TSTNLSQSSASQPPRQOQGVKLTMAQLQLQIUGHGNGLVY-----VIOGQGO---TTGQ 2039
Db 2761 TTYVFOASKIEP-----IVDLRAGVVLNNTGASSAGTFILMAASVLQOQOTQDDPTTQ 2814
Oy 2040 LQILPQGVVLPQPGQQLQMAAMPNGTVQRFLETPPLATATTASTTTVYSTAAGTGEQ 2099
Db 2815 I-----LQMANFO-FQSVPTSSGASTMDTSPVMVTA----- 2846
Oy 2100 ROSKLSPOQOVHODKTLPPAQS---SSVGRPAKAPQ-----TAQSAARPQOTOPQSPAQ 2151
Db 2847 ---KIPPVQIKRTMAQAKAGISGVKVPQPPQVNVKVLPTISIVTQSOQVQKNSNLQ 2903
Oy 2152 PEVQTPQEVQOTQTVSSNHPSA-----QPTHQSSKROY----- 2186
Db 2904 SQVKGAAAGSGTGTGCA-PPSIAASKPLQKTKNMIRPIHLEVPKVMKTPKYQONHSL 2962
Oy 2187 -AAQSPQSNVQSGSPVRV---QSP-----SQTIRIPST-----PSQLSPQOQS 2226
Db 2963 LQOQOQOQOQLOQOQIPAVVNOVPRKXTISQQRIPATQOQOQLOQOAMINIPQOQOPLQOQ 3022
Oy 2227 QVQTTTQOP---PIOPHNSLQIPSOQOPQOSPOVQOSSQTGLSSGOTLNOVSVSPSR-- 2281
Db 3023 QVQVQSPMTITLAEAPVVOQFVEPQALEQOEIANRVQHFSTSSSSSSSSCSLPTNVV 3082
Oy 2282 -----POLQIQOPQ-----QVIANPQLOQOVQ---VLSQIOS 2311
Db 3083 NPMQQAAPSTSSSTTRPTNRVLPQOQROEPARLNSCEPQVSSPTPPKVEOPTIHQMTS 3142
Oy 2312 QVVAQLOAQOQSGVPOQIQLPQIQOQSSAVQTHQIQNVTVVQAASVQEQLOQHVQQLRQ 2371
Db 3143 ASVSKCYAQKSTLPSRV---YEAEIKVSSVLES---IVPDVTMA---ILEQRPQISYITE 3195
Oy 2372 QQKKQOQOQIEIKREHTLQASNOSEITQOQVVMKNAVIEHLQKKSM-TPAREEN--- 2426
Db 3196 GLYEXNSPGEKTEQQLLOQOQREQLNQOLVNNGYLLDKHTFQVEMDFDVYREEDLEE 3255
Oy 2427 -----QRMIVCN-----QVMKYLDKIDKE----- 2447
Db 3256 EDEDDDFSLKMTSACNDHMSDEBPAYVKDISKILDNLTJNDCADSIATATTMEVDAS 3315
Oy 2448 -----KQAAKKRRRESVEQKRSKQNAUTKLSALLFKHKEQLRAELIKRA 2492
Db 3316 AGVOQAVEVLLATTAQASAPTEFEQALETAVAEAATYINEMADN----- 3362
Oy 2493 LLD-KDLQIEVQBELKRDIKKKEKDLQLAQATAVAAPCPVTPVLPAPRAPPSPPP 2551
Db 3363 VLDLQQLQNGVELELRR---KEQKRTVSEQEQSAAIATP---TAAAPERQOIQEPKKM 3417
Oy 2552 PG 2553
Db 3418 TG 3419

```

RESULT 7
 T34434
 hypothetical protein K06A9.1a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T34434
 R:Geisel, C.; Gattung, S.
 Submitted to the EMBL Data Library, December 1996
 A:Description: The sequence of C. elegans cosmid K06A9.
 A:Reference number: 221525


```
0Y 1037 ETKCPKNSIENDIEKYSDLASRGOEPTKSKTKGNDIFIDDSKILASADIGTLICKNKK 1096
Db 1941 EKKOPREKIE--IDENSSSSNSGOE--KPTTKG----- 1971
0Y 1097 PLIOESDPTVSSSKSALH--SSVPEKSTNDRDATPLSRAMEFGKLGDCSESNTLENS 1154
Db 1972 -IVSSSTSAISSSTIAEPHVTTISISTTKDMT-----SSKSPENVWSSS 2018
0Y 1155 DTVSIODS---SEDMIVONSNEISIEQFRTREDOEVLPLKELVSGESTGNCEDRL 1210
Db 2019 PEVSTSSSKSTIASSETVSTPSESSSE-----APLTSPPAT----- 2056
0Y 1211 PVKGTAEANCKKRSQOKKLEFPVYKCSODIKLNTDKANNRESEKGCORTSTFOING 1270
Db 2057 -----TTEVITESSVSKSTPKKESSSEITVKKLSKSS----- 2087
0Y 1271 KDNKPKIYLKGECKEISRSRVASGVNEPV--NNINKLIIPENDIKSLVKSAIRPFIN 1328
Db 2088 -----PEVTSVSSKSSPTSTTSOSVSTVETSKSTVLSSEAPVSTSP 2133
0Y 1329 GDVIMEDFERNRSSETKSHLLSSDAEGNYRDSLETLPTSKESDSTQTT----- 1377
Db 2134 TEV-----HTSSEKPKLSASSTT---GDNSTPTSTSLASVKSTSAEGTSASVA 2182
0Y 1378 -----TPSACSPESNSVNOVEDMEIETSEVKATTSITSEESNLNDPIDENGLPI 1430
Db 2183 PVKLSLSLSDVQSPSTKTFDATESSTVQASETSSTGVSSTSPESHVTKLSTSS--- 2238
0Y 1431 NKNENVNGESKSKRTVTEVTMTSTVATESKTVIKKEKDKQTVSSTENCAKST---V 1486
Db 2239 NPSSTVPPTSPKSTPVPSTBOPTSTTPSGSLTPMNSNEVLTTSEHVSSSLSPV 2298
0Y 1487 TTTTTVTKLS-----TPSTGSVDIISVKEOSKTVVTTTVDLTTTGLTVTSVTS 1540
Db 2299 SOGSTTPNNLSSTSVTEPKTSEVSLNEEPTTEAPTLSPDLISTTNNLSOSSTV- 2357
0Y 1541 KEVSTRDKVKKLKFSPKPTRSGTLPSTRKTVYTKSKSIVLNPDLKKLARKGCIRE 1600
Db 2358 ---STEDRSEISENSEKPTSAPELVTSSVTHVASSSPD---VPTSESEPDLTSGSTE 2410
0Y 1601 -VPYFNYNKPALDIMPV-----SPRPEGITWRRLQTVKSLAGVSLMLRLMALSRLW 1654
Db 2411 NIP--EASKKQIISTPTPTDITTAEEPTKSTMSFDSLSTSNVSES----- 2456
0Y 1655 DDMAAKVPGGGSTRFTESETE---ITTELIKRRDVGPGYIRFEYCIRKIICPIGVE 1710
Db 2457 -----STPPESSKSPVSSSTEGISVISTEFSK-----VEE 2488
0Y 1711 TPKEPTTPQKRLSSALRPKRPETPKOTGVIIETWAEELLE-----LMEI 1758
Db 2489 S-----TISVLEEDLTKT--TPSPILLEETTTASETSEPLETDSLTVSVRIHEL 2535
0Y 1759 RAFAEREKEKAOVEQOKKRLBOOKPVIATSTTSPTSSTSTTSAPKQWVAPATSS 1818
Db 2536 TTSSSENVPE--SESTTSSSEKSPSOEPAGILITVVPVTSVSLTJASE-----IEA 2587
0Y 1819 VTTGTRMVLITKVGSPATVFOONKNEHOTFATWVKVQOSNSGVVQVQOKVLGIIIPSTG 1878
Db 2588 ITSNNPF-----KQGR-----PITTSKPSLVKST-- 2612
0Y 1879 TSOQFTSFQPTATVITRPNTSGSGGTTSSNSOVITPOIRPGMVIKRPLOOSTLGKAI 1938
Db 2613 TSPSTVTSSEPESTR-----TTVSTTVST-----TPTTEETTSSESL 2651
0Y 1939 ITTPVWVGAGQAVMTOLIRG--QPVSTAVSAPNTVSTPGOKSLTASTSTNSIOSSAS 1996
Db 2652 IIT---AAPSKESTTESSEAPTPPAKTSKTPSNVSS-----TSKSTENETSTS 2701
0Y 1997 QPPRPOQGVKLTLMAQLTQTOGHGNGCLVTVIOGQGTGOLQLOPQGVVLPGPGQO 2056
Db 2702 Q-----SGSLE----- 2707
0Y 2057 LMOAAMPNGTVORFLFTPLATTATTATTTTTV-STTAACTGEOBKSLS---POMQVHO 2112
```

```
Db 2708 --SSTMSSTSSPEPINAIPVTVSSSEASTTLEENSTSSPTSSSEASVKLSLPEPSTSE 2765
0Y 2113 DKLTPPAOSSVGPAAKAPQTAQAPSAPOQIOPQSPAPAEVOTQPEVOTQTVSSHVS 2172
Db 2766 ATYV-----SSNAPAI-----TMSSEHRELSTVSSSPSPPELPLSTVSPNNVTASTIS 2817
0Y 2173 EAQPHAGSSKQVAAQSP--QSNVQGSPPVRVQSPQRRIRPSTPSQL-----SFGQOS 2226
Db 2818 E-----EPILSVTSSSTPRV-----RLITGTPDDLIVSVTPSHCN 2854
0Y 2227 QVQTTSPPIPIQPHISLQIPQSQPQSPQVQVQSSQTQLSSQTLQNVQVSSPSRQLOI 2286
Db 2855 RQONITVASSVPNSNSTSPILPSESLTTPQP-PPTTTAKPATTSKGRPPSIQPAEMF 2913
0Y 2287 QOPOP 2291
Db 2914 TTPAP 2918

RESULT 9
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence-revision 03-May-1994 #text-change 07-May-1999
C:Accession: B48666
R:Schueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCCH>
A:Cross-references: EMBL:X65551
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 412; DB 2; Length 2897;
Best local similarity 18.5%; Pred. No. 1.4e-07;
Matches 551; Conservative 391; Mismatches 1062; Indels 980; Gaps 137;

0Y 291 OCEVCAHHVPGVTDCAETOKKPYIRH-----EPITG-----YDR 326
Db 34 ECDIRQLQLVVSKQKQKIEIHDEALIHNFSSNPTQVNGSVIDEVYRLKHGVITIDR 93
0Y 327 SRKKYFELNRLRIIEEDTENENEKKIWIYSTVYQVLAELIDCLDKDYWEALCKILEMRE 386
Db 94 SFR-----YENESLQNGRKSTEPKRIREQEPARVRSRSFSSD---PDEKAQ 138
0Y 387 EIHRRHMDITEDLTKRKGNSKSLAANEILESTAKKGIDN-----VKSPEETED 440
Db 139 DSKVSKITE---GKVSQNPLEF---DENLPPNPLKGEAPTKRKSILVMHTPPYLKRI 191
0Y 441 KNETENDSKDAEKNR---EFEDQSL---EKSDQDKT-DDEDEQKSEVGFKESEKN 492
Db 192 IKEQPPQSKQESGSEIHVEKAGSLVSPRAPSPKRTVVASQRRRSCKTAPASSSKQ 251
0Y 493 GELSESPGAKGAGSGSTRITRLRNPDS-----KLSQKSOQVAAAHANKLEKEGE 546
Db 252 TEVPK-----RGGERVATCLOKRVISRSQHDILQMGSKRRSGASEAN----- 295
0Y 547 VLVVNGGELISLSTPK---EVIKMG---NINNYFLQGEQKRYV--YANQSTN----- 593
Db 296 LIVAKSMADVNVKLGAKQOTQKVIKHPQPSMNKRQRRPATPKKPVGEVVSQFSTGHANSP 355
0Y 594 -SFALNKHQRHEDHKRR--HLAHKFLTPAGEFK-----WNGSVHGKSVLTISTL 641
Db 356 CIIITIKATEKVHVARRYRLNNEPISQMKDFKEDLSGIAMFKTPYKEDPOL-STRC 414
0Y 642 RLITQLENNIPSSP-----LHPNWSHRAMWIKAVQMSCK--PREFALALAILCEA 691
```

Db 415 HIAISSENILCKQFOGOTDSGCEPFLPTSESFCGNVEFSSQNAKQPSD-----KCS 466
Qy 692 VKRYVWMLPTWREFLGH-----RLHMTSLIERE-----EK 721
Db 467 ASP-----PLKROQIRBNGVNAKTPRNTYKMTSLTETSDTEPEPSKTSTVSNRSGSTEE 522
Qy 722 EKKK--KEKOESEETM-----OATWYKTYFPVKHQWKOKEEYRVTVGSGWS 770
Db 523 RNIOKLPVEKSEETTELVEICILKRGOKATLLOQ-----RDEGKEI-----566
Qy 771 WISKTHVREPRPLPGNTVNVKRSLEGTKNNDENMDESCK-----RKCS--RSPKIK 823
Db 567 -----ERPEET--YKENITELKEN--DEKKAMKRRSRTGQCAPMSDLTLDK 609
Qy 824 IEPDSE--KDEVKG-----SDAKGADQNMEMISKITEKODVVELLSDSDKCKE--874
Db 610 SLDPDTLMKJTAGOMLLOTQDHAKAPKSEKGITKM-----PCQSL 651
Qy 875 EPMEVDDMKTESHVNCQESSQDVVNVSE-----GFHLRTS-----911
Db 652 QP-----EPITPTHTKOQLKASLGKVGKELLAVGKFTTSGEHTHTHREPAQDGKSTR 707
Qy 912 -YKKTK-----SSKLDGLLEIRIKQFTLEKORLEKIGKIGIKTSTNKKLSE 965
Db 708 TFKESPQOILDPARVYGM--KKMPTRPKKEAQSLEDL--AGFKELFQTPGSESMTD 762
Qy 966 SPVITAKKECCOSDMROBESPANNDQPE--DLIOGCSQSDSVLNMSPSHNTNLKPK 1024
Db 763 EKT--TKACKSPPEPSVDPTPTSTKOWPKRSJRKADVEEFLAKLKLPPSAGKAMLTPK 819
Qy 1025 DRYLDDVSI-----RSPETKCPKNSIEN-----DIEKVS 1055
Db 820 PAGEDEDIAFMGTPOKLDLAGTLPGSKROQLOTPEKQOALEDLAGFKELFQTPGHT 879
Qy 1056 DLASRQ-----EPYKSKTKGNDPFLDSDKSLASADIGTIL--CKNKKPLI 1099
Db 880 ELVAAGKTKTKIPCDSPQSDPVDPTPTSTKOR-----PKRSIRKADVEGELLACRNLMF--931
Qy 1100 QESDPTVSSKSAIHSVVKSTNDP-----APPLSRAMPFGKIGCSEDSSTLNNSS 1154
Db 932 -----SAGKAMHTPKSVGBEKDIIIFVGTPOVL-----DLTENLT 968
Qy 1155 DTVASIODSEEDMIYONSNESISEOFTRBQDVEVLEPLKCELVSGESTG--NCEDRLLPVK 1213
Db 969 GSKRRPQTPKEAQALEDLGLFKELFQTPGHTBEA-----VAAGTKTKMPCSSPPES 1021
Qy 1214 GTEANGKKPSQOKKLEPRPVNKCSDQIKLNTTDKKNENRESEKQGTSTQJINGKDN 1273
Db 1022 ADPTSTRROPKTPLEKRDYOKELSAK-----KLTQTSGETTHTDKVPQGED 1069
Qy 1274 KPRIYLKGECLKEISESRVYSGNV--EPKVNNIKIIPENDIKSLYVESAIIRPFINDVI 1332
Db 1070 KSIJNAFETKQKLDPAASVTSGRHRPKTK--EKAQPLEDLAGM--KELFQTP-----V 1119
Qy 1333 MEDFENERSSETSHLSSSDAAGNVYDSLETLPSTKESDSTQTTTPSASCPSNSVNOY 1392
Db 1120 CTOKPTHEKTKIACKSQD-----PYDPTPSKKPQSKSLRK 1158
Qy 1393 EDMEIETSEVKKVYTS-----PITSEESNLMDPLIDENGLPINK--NENVNGES 1440
Db 1159 VDVEEEFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPQOKLDLTENLGLSK 1213
Qy 1441 KRRTVITEVTMTSTYA-----TESKTVIKVEGDKQTVYVSTENCAKSTVTT 1488
Db 1214 RRLQTPREKAQALEDLAGFKELFQTRGHTESMT--NDRKTAVACKSSOPDLDKMPASS 1270
Qy 1489 TTTTVTLTSTPGSVDIIVSKEOSKTVTTVTVDLSLTGTGGLVTSMTVKEYSTRDK 1548
Db 1271 KRRLKTSLG--KVGVEELLAVGKLTQTSGETTHTHTPELPGDKSKMAFESRQOILDSA 1328
Qy 1549 VKLMKESRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPPNDL--1589

Db 1329 ASLTGSKROLRTPKKSEVEPDLAGFIELFQTPSHTKESMTNEKTKVSYRASQDPLVDT 1388
Qy 1590 -----KKIARKGIRVFPVFNNAKPALDIWPPSPRPFGITWVRLOTVSLAVY 1641
Db 1389 PTSSKQPKRSLRKADTEE-----EFLAFKQTPSAG--KAMTPKPAVE 1432
Qy 1642 SLMLR--LTMASLWMDMAKVPBGGSRTETSEETITTEILKRRDVPYGR-----1694
Db 1433 ENDINFLGTPVOKLQDQPNLP--GSNRLQTRKEKAQALEELT-----GFRELFPQTR 1483
Qy 1695 -----FEVCIRKILC--PIGVP--ETPKETTPPOKGLRSS-----ALPKRP-----1733
Db 1484 CTDNPTADKTKTKKILCKSPQSDPADTPNTKORPKRSJRKADVEEFLAFKRLPSACK 1543
Qy 1734 --ETPKQTPVILIEFWAESELELMEIRAF--ARVE-----KEKAQ 1772
Db 1544 AMHTPKAA-----VGEK-----DINTFVGTPVKLDDLGNLBSKRRPQTPKEKAA 1591
Qy 1773 VEOAKRRLQOKPVTIATSTTSPTSSTSTISPAQKVAVAPISGVTGTGKMLVTTKYG 1832
Db 1592 LEDLAGFKELFQTPGHTESMTDDKITEVCSKSPQDPVKTP-----TSKORLEKISLG 1645
Qy 1833 SPATVTFQONKNPHQTFATWVAKOGSNSGVVOYQKVL--GIIPSTGTSQOTFTSFOQR 1890
Db 1646 K-----VGKVEEVLVGLKLTQTSKTTQT-----KKAQ 1669
Qy 1891 TATVTIRPNTSGSGT-----TSNSQVI-----TSPQIRPGMTVTRTP-----LOQST 1933
Db 1670 -----HRETAGDKSIRKAFKESAKOMLDPANTGEMWP-----RTKKEAQSLEDLA 1718
Qy 1934 LGRATIRTPVWQPGAPQOVMTQIIRGQPVSTAVSAPNTVSTSGOKSLTSATSNIOS 1993
Db 1719 GFELFQTPDHTHEESTDDKTKI-----ACKSPPESMDTPTST-----1758
Qy 1994 SASQPPRPOQGVKLL--NAQLTQLTQHGNGNGLVAVIOGQQTGQLOLIP--OGV 2047
Db 1759 --RRRKTPDKGRDIVEEISALKQLOQ-----TTHDKKPGDEDKGI 1798
Qy 2048 TVLPGGQOIL-----QAAMPNGTVQ-----RELF--TPALATATTASTTTT 2087
Db 1799 NVFRETRAKQKLDPAASVTSKRPRTPKGAQPLEDLAGKELFQTPVCTDKPTHEKTT 1858
Qy 2088 TVSTTA-----AGTGE--OROSKLS--POMOVHODKTLPPAOSSSVGPAAQOTA-----2134
Db 1859 KIACRSPQDPDVGTPITFKQSKRSJRKADVEEESLALRKRPVSGKADTPKPAQDGK 1918
Qy 2135 -----QPSARPOQTOPQSPAQPEVQOPEVQOTPVYS-----HVPSEAQ 2175
Db 1919 DMKAFMGTPVOKLDPGNLPGSKRMVQTPKE-----KAQALEDLAGFKELFQTPGTRK 1971
Qy 2176 PTHAQSCKPOVAASQPOSNVQOSQSPVAVQSPQTRIRPS-----TP 2217
Db 1972 PT--TDEKTKIACKS--PQ-----PDPVDTPASTQORPKRNLKADVEEFLALKKRP 2022
Qy 2218 S-----QISPQGOQOVQTTTSPRIPQHTSL--QIP--SOGQPOQOP-----2256
Db 2023 SAGKAMDTPKPAVSDENKINTFEVETPVOKLDLGNLPGSKROPQTPKEKAQALEDLVGRK 2082
Qy 2257 ---QVOSSQTOLSSGQTLINQVSVSSPRQLOQOQOQVYIANP---QLOQOQVYLSOI 2309
Db 2083 ELFPQTPGHTHEESMTDKITEVCSKSPQESFKTSRSKORLKLPLVVKDMKEPLPAVSKL 2142
Qy 2310 --OSQVAAQLOAQOSGVPQOIKL--OLPLOIQOQSAVQTHQIONVYT--VOAASVCEQOLR 2364
Db 2143 TRTSGETTQTHTEPTGDSKSIKAFKESPKOILDPAAVTSKROLTRKEKAQALEDLVD 2202
Qy 2365 VQOL-----RDQOK-----KQOQOIEIKR-----EH 2386
Db 2203 FKELFAPRGHTESMTIDKTKIIPCKSPPELDTATSTKROCKTPRREVVEKELSAVER 2262
Qy 2387 TLOASQOSETIOKOVVKNHNAVTEHLKO--KSMSTPAERENOMATVYCNQWAKYIILDKID 2444
Db 2263 LTQTSQSTHTHREKSPASGEG--IKVLKORAKKKNPNEVEEPPSR-----2305

Db 926 -----ERPEET--KENELKEN--DEKMAKMSRTWGOCAMSDL 964
 Qy 820 KKIIEPDSE--KDEYKG-----SDAKGADONEMDISKITEKKDQVKELLSDSDKP 871
 Db 965 TDKSLPDELMDQARGOQLQTODHAKAPKSEKGIITKM-----P 1006
 Qy 872 CRE-EPHEVDDMKTESHVWCSSSOVDVNVSE-----GFHLRTS----- 911
 Db 1007 COSLOP-----EPINTPTHTKQOLKASIGKVGKEELLAVGKFTRTISGETTHHREPADG 1062
 Qy 912 -----YKKKTR-----SSKIDGLLEIRIKOFTLEEKORLEKIKLEGISKIGTSTNSK 961
 Db 1063 KSJIRTKRESKOQLLDAARATGM--KKMPTPKHEAQSLEDL---AGFKLFQTPGSE 1117
 Qy 962 NLSSEPVITAKGECOSDSKROQSPNANDPE-DLIOGCSOSDSVLRMSPSHTTK 1020
 Db 1118 SMTDEKT---TKIACKSPPEVDTPTSTKOWPKRSLRKADVEEFLAKLTPSAGKAM 1174
 Qy 1021 LYPKDRVLDOVSI-----RSPETKCPKONSLEN-----DIE 1051
 Db 1175 LTRPAGGDEKDKAFKMGTVOKLIDLATLPSSKROLQTRKKAQALEDLAGEKLEFQTP 1234
 Qy 1052 EKVSIDLASRGQ-----EPTKSKTKGNDEFIDSKLASADDIGTLI--CKNK 1095
 Db 1235 GHPEELVAAGKTKTKIPCDSPQSDPVDPTSTKOR-----PKRSIRKADVEGELLACRNL 1288
 Qy 1096 KPLIOESDPTVSSSKALMSVPKSTNDRD-----ATPLSRAMDEPGKLGCSSESSTL 1150
 Db 1289 MP-----SAGKAMHTPKPSVGEKDIIFVGTPOKL-----DLT 1323
 Qy 1151 ENSSDTVSIOSSSEDMIVONSNEISIEOPRTREQDVEVEPLKCELVSESTG-NCEDR 1209
 Db 1324 ENLTGSKRRQOTPREKQALEDLGFEKLEFQTHPEEA-----VAAKTKMKCESS 1376
 Qy 1210 LPVKGTEANGKBSQOKLEERPYNKCSDOIKLKNTTODKNNRESEKKGORTSTFOIN 1269
 Db 1377 PPSADPTSTSTRQPKTPLEKRDOKELSAK-----KLQOTGSETHTDQVP 1424
 Qy 1270 GKOKPKIYIKGELKESISRVSGNV--BPKVNNIKIIPENDIKSLVYKESAIRFIN 1328
 Db 1425 GGECKSTINAFRETKADLPASVYTGSKRHPKTK--EKAOPLEDLAGW--KELEFQTP-- 1477
 Qy 1329 GDVIMEDPNERNSSETKSHLSSSDAGNYSLETLPTSKESDSTOTTTPSACPESSNS 1388
 Db 1478 ---VCTDKPTTHEKTKIACRSQD-----PVDPTPSKPKSKR 1513
 Qy 1389 VNOVEDMEIETSEVKVYTS-----PITSEESNLNDPIDENGFLINK---NENV 1436
 Db 1514 SLRKVDVEEEFFALRKRTSPSGKAMHTPKPAVSGEKNIVA--FY--GTEVOKLIDLENTL 1568
 Qy 1437 NGESKRTKTVITEVTMTSTVA-----TESKTVIKVEGDKQOTVAVSTENCAS 1484
 Db 1569 TGSRRRLQTRPEKAQALEDLAGEKLEFQTRGHNEESMT---NDTKAVACKSSOPDIDKN 1625
 Qy 1485 TTTTTLTVTKLSTPTSGSVDIISVKEOSKTVTTVTLSLTGTLVTSMTVSKEYS 1544
 Db 1526 PASKRRLKTSIG--KVGVKELLAVGKLQTSGETTHTEPTGDSKMAKAFESPQOI 1683
 Qy 1545 TRDVKLMLKTSRPKTKTSGA-----LPSTYK--FVYKSTKKSIFVLPND 1588
 Db 1684 LDSAASLTGSKROLRTPKGSEVEDLAGFLEFQTPSHRKESMTNKTTKVSTRASOPD 1743
 Qy 1589 L-----KTLARKGIREVPYFNVANAKPALDIMPSPPTGTIMRYRLQTVKS 1637
 Db 1744 LVDTPTSSKPOPKSLKADTE-----EFLARKQTPSAG--KAMHTPKP 1787
 Qy 1638 LAGVSLMLR--LWASLRWMDMAAKVPPGGSSTRTSETITTTTETIIRKRDVCPYGR-- 1694
 Db 1788 AVGEKENDINFLTGPVKLODQPNLP--GSNRRLQTRKKEKAQALEELT-----GPREL 1838
 Qy 1695 -----FECIKRITC--PIGVP--EPPKTEPTTPORUGLASS-----ALRKRP 1733

Db 1839 FQTPCTDNPJADEKTKTKILCKSPQSDPADTPNTKQRPKRSILKRDVEEELAEFRKLTP 1898
 Qy 1734 -----EPPKQGPVITIEFWAVEEELMEIRAF-----AERE-----KE 1768
 Db 1899 SAGKAMHTPKRA-------VEBER--DINTFGTPEYKEDLIGNLPKSGRRPQTRKE 1946
 Qy 1769 KAOAVEQAKRLEQKRPVIAVTSSTPSSTSTSPSAOKVMVAPISGSVTTGKMVL 1828
 Db 1947 KAKALEDLAGEFLEFQTPGHTEESMTDKITEVSKSPQDPVKTP-----TSKORLK 2000
 Qy 1829 TVGSPATVTPQONKNFHOTEPATVWKQOGSNGSVYOQVKV--GIIPESITSOQTPFS 1886
 Db 2001 ISLKG-----GVKKEEVLVPVKLQTSKTTQTP-- 2028
 Qy 1887 FQPRFATVTRPNTSGSGT-----TSNSQVI-----TGPOIRPGMTVIRTP--L 1929
 Db 2029 -----HRETADGKSIAFKESAQOMLDPANYGEMERW-----RKPKEAOSL 2073
 Qy 1930 QOSTLKAIRTPVWVOPGAPOQVMTQITRGQVSTAVSAPMTVSTPQKSLTSATSTS 1989
 Db 2074 EDLAGKLEFQTPDHTEESTDDKTKI-----ACKSPESMDPTSTP- 2117
 Qy 1990 NIOSSASOPRRPOQGVKLT--MAOLTOLOGHGNGLTVIIOGGQTTGLOLILP-- 2044
 Db 2118 -----RRRPKTPPLGKRDVIEELSAKLOLQ-----THTKVPQDE 2153
 Qy 2045 -OGVTVLPGGOOLM-----QAAMPNGTVQ-----REL--TPLATTATATAS 2083
 Db 2154 DKGINVFRETAKOKLDPASVYTGSKRQRPTRPKKAQPLEDLAGEKLEFQTPCTOKPPTH 2213
 Qy 2084 TTTTIVSTTA---AGTEB--ORQSKLS--POMQVHODKTLTPPAOSSVIGPKAOPQTA- 2134
 Db 2214 EKTTKIACHSPQDPDPVGTPTIFPKQSKRSLRKADVEEESIALRKRPSSGKAMDTPKPG 2273
 Qy 2135 -----QPSARPOQOTQOSPAPQEVQTOPEVQOTQTVSS-----HVP 2171
 Db 2274 GDEKDKAMFNGTPVOKLDPGLNLPSSKRPQTRKE-----KAQLEDLAGEKLEFQTP 2326
 Qy 2172 SEAQPTHAOSSKPOVAOSPOSNVQOSBPVAVQSPQTRIRPS----- 2215
 Db 2327 GTDKPPI--TDEKTKTKIACKS--PO-----PDVDTGASTKORPKRLRKADVEEELALR 2377
 Qy 2216 --TPS-----QLSPQOQSOVOTTQSOPPIQPHISL--QIP--SQGPOQOP- 2256
 Db 2378 KRTPSAGKAMDTPKPAVQSEKNINTEVEVYOKLIDLGLNLPSSKRPQTRPEKAEALEDL 2437
 Qy 2257 -----OVQSTQTLSSGOTLNQVSVSPSRQOLOOPOPOVIAVP-----OLOQOVQV 2305
 Db 2438 VGEKLEFQTPGHTEESMTDKITEVSCSKSPQSPESFTSSSKORLKIPLVKYDMKEPLA 2497
 Qy 2306 LSOI--OSOVVAQIOAQOSGVPQOIKL--QLPLOIQOQSSAVQTHQIONVVT--VOAASVOE 2360
 Db 2498 VSKLTRTSGETTQTHNEPTDSSKIKAFKESPKOILDPASVYGSRRQLTRKREKARALE 2557
 Qy 2361 QLOROVOL-----RDOQOK-----KKOQOIEIR-- 2384
 Db 2558 DLVDFKELFSPAGHTEESMTIDKNTKIPCKSPPELDTATSTKCPKTRPREVEEELS 2617
 Qy 2385 --EHTLOASNOSEIIOKOVYMKHNAVIEHLQ--KKSMTPAEREENQMIIVCNOVKYLL 2440
 Db 2618 AVERLQTSOQSTHTHKREPSGDBG--IKVIAKQAKKPNVVEEESRR----- 2664
 Qy 2441 DKIDKEQAKKRRKEESVQOKRSSQONATKISALLFKHKEOLRA-----EI 2487
 Db 2665 -----RPRAPKEKAQPLEDLAGEFLESETSGHTQESLTAQKATKIPCESPPELV 2713
 Qy 2488 LKRRALLDKLOLEVOELKROLKIKKENDMLAQ----- 2523
 Db 2714 VDTTASTKRLKLRVQV-----VOVKEEPSAVAFQTSGETTADKPEAGDEGITALKE 2768
 Qy 2524 -ATAVAAPCPVPVLPAPAPPSP--PPGVOHTGLSTPLTPVYASQKRRR 2574
 Db 2769 SAKOTAPASVYGSRRRRAPAPRESAQALEDLAGFDPDPAAGHTEESMTDOKTKIKICKSS 2828

QY	1338	ERNSEIETHLLSSDAEENYADSLLETLPBTKESHSTOQTTPSASCENSNVQVDEMI	1397
Db	1484	EQSDSESTAFKIRDDPAQ-----KTEELISKODDEKTYT-----DCKPRKPEDSEA	1531
QY	1398	ETSE--VKKYT--SSPTISE-----EBSNLNDPIDEENGLPINKNEVNGSKRRKYTVEY	1449
Db	1532	TPKKRYVKKKTKQSDSADADSLADYSKSLSDVEEKPKKVKLYKKT-----EKSDSYISE-	1586
QY	1450	TYMTSVIATSKVIVKEGDK-----QIVSSTENCAR-SVYTTTTYVVKLSY	1498
Db	1587	TSSVDITIKPES-VEIPLTEAKEDMILHNRESTDSAVESEPKNAAHDJIEKTTDMMTKRS	1645
QY	1499	PTSGVSDIISVSEOSKTVVTT-----VDSLTFTTGGT-----LYVS	1536
Db	1646	SA-----ITSDDQSISSKTSEGRRRRRRTGFAKSPASOTLALRODNEIEAELLAD	1699
QY	1537	MTVSEKYSTDK-----VKLMFSRPKTRSGTALDSYKREYVTKSKSIY	1583
Db	1700	DTVMKNGKADLNSRCEHMSHTPEFRTLIDIVEPTDSGEIYATGCTESHETILKVEE	1759
QY	1584	LPNDLKLKARKGIR-----VPFNNAKPAIDIMPYSPREPTGIGTRRYLQTVKS	1637
Db	1760	LPVPEVYLLPFRKTSKGQGEVTISVTLNPIIDISKVW-LADGAP-LEINKDYSIDTVG-	1816
QY	1638	LAGVSLMLRLTMASTLBMDDMAA-KVBPGGSGSTR-----ETSETEIYTT	1680
Db	1817	-CNSVLTLR-----RAKEXDGGKRYKVCDDGDCSTHLSIOGKPVLKNSETPYVTVDKDD	1871
QY	1681	-EIIKRDQPY-----GIREYCIKRIKICPI-----GVPEPKK--EIPTP	1718
Db	1872	QFSLIIVAYDSNPASFMSMTVDGKLEDFDRSRIRIDVDDGLKTRKGSKTDAGEYKKL	1931
QY	1719	CRKGLRSALTPKRPPEPKOTGVIIE-----TWAEETEELMEIRAFABEREK	1769
Db	1932	NEFGVNAKFPVKNPIKNDIPSAPOGVSVAKAESODLHIEMTAPTEDNGAEVTSYV--EKEE	1989
QY	1770	A-----QAVEQOAKKR-----LEOQKPVIAITSTSP-----TJSSTSTISPO--	1808
Db	1990	SCRKRFHVATVNGKRTSYVDDLEIETPYIVARIAVKNFEGTEIERKPVOTSSPOVP	2049
QY	1809	KVMVAPRIGSVYT-----GTRKMLTKRVG	1832
Db	2050	TYEFPPTIDANTSYSCSLSNPKPIEDOGSPVYGVYDKRENKGEMQNMKGELVETSEFN	2109
QY	1833	SPATVTFQONKNFHOATVWKGQ-OSNSGVVOQOKVLGITPS-----STGS	1880
Db	2110	VBA--LSGKEYEFKIEACAGELBNSNVWSKLTVEGLVETIILDMPVKVLNDKY	2166
QY	1881	QOTFSRPRATVYTIIPNTSG-----SGCTSNQ-----VITG-----	1915
Db	2167	EVTWMSDGEKEFVVOYKSDGSSIWASVDIGCPRSESAATSKIIDGLBEGIPYFRNAAR	2226
QY	1916	-----POIRPGMTVIRTPLOOSTL-----G	1935
Db	2227	NQHGGERSEPTIPVVLADAPRVILKAIKPVKIPKKGELRTECHAAGHPAHEYIWKDG	2286
QY	1936	KAIIRT--PVMOPGAPQOVMIOIINGOVST-AVSANTVYSSRTPQOKSLTASSTNSNI	1991
Db	2287	KEIIPIDENTELIWNKGSMSALIHILAGEVGLYKVLVENIHGTASESEVGISVRAHF	2346
QY	1992	OSSASOPRPOOG-OVKLT-----MAQLTOL-----	2016
Db	2347	NSSPSELIEIEGHDIELTCVSDSEAEVVMYWDKKIVASDRVGFVMAARKRLIRIGS	2406
QY	2017	-----TOGHGNGGLIYV-----IGOGOT-----TGQ	2039
Db	2407	TDADSGVYKCTETDGRSRTEGEVIVNQDEPHILLVQDAILVKDFETMYLFCETSKPVK	2466
QY	2040	LOIIOGVYIYJPGEQOLOAMAMPNGVORPLFTPLATVATASJTUTTVYSTAAGTGEQ	2099
Db	2467	VMEFKNGEIVQONKKAIMEWDGRKATILIEKTFDKHDIGATYTSVSEKRTSNAPALIVEY	2526

```

OY 2100 ROSKISPO-----NQVHDKTLPPQAQSSVGPAAKAQQTQAPSRPQOPQOPQOPAPQEV 2154
Db 2527 APNLIIPETRIDGVVHAAGNEPFA-----VFSGFPIPTIHLTNNGPFLK 2522
OY 2155 QTOPEVQQTQTVSSHVPEAPQPTHAQSSKPPVAQASQPPQSNVQASQSPVRQSPQRIIP 2214
Db 2573 AIAVATEXDDSVSVAM--KQVTLDNSGTVHVAES--PLGQCIKEIPLKIID-----KP 2622
OY 2215 STPSQLSFGQSQSVQVTTTSQPIPIQPHTSLOIPQGGQPPQSQPVQSSQTLTSSGQTLNOV 2274
Db 2623 SAPCCL-----QFKEVTEDSVFLSWQP---PLETNGAP-----LIGVYIERK 2661
OY 2275 SVSSPS--PPQLQIQPOP-----GVIAVPPQLQQQ-----VOVLSIQISQV 2313
Db 2662 AVDNRMKRPCCGVKTKTLFEVAEDLCNQVYGFRILAVNEVGESEPCDVTVLTTLESSEP 2721
OY 2314 VAOIQAGSGVPPQIKLQLPPIQIQOQSSAVQTHQIQNVYTVQAASVQE-QLOLRQQLR--- 2369
Db 2722 VS--ESELTEPFIKILKTP--QVTVAVD--ETKVTLMEECPETSILKVEKRVGD 2772
OY 2370-----DQQKKRQOQOIEKREHTLO-----ASNQSEITIQOVMAKHNVAIEHLK 2413
Db 2773 SQWLEIANTDRKKFKDRSLTESGEYVQVATGHAIVSSPEEENPVPKILVPGSEMASK 2832
OY 2414 QKSMTPAREENQRMIVCNOQMYKYLIDKIDKEEQAAKRRKREBSVEQKSKONATKLS 2473
Db 2833 TEKTDAAKSESEQKS--AEIIV-----AEQOVQOQASQESTVEAVEEKKTK----- 2877
OY 2474 ALLFKIKEDRALILLKRRLLDKLDQI--EVOEELKRLUKIKREKIDMLQLAQATAVAAPC 2531
Db 2878-----KVVKKKAENKGETTLOEYAEKLLKKAIVAEKVQDESRBSLQASS--- 2922
OY 2532 PPVTPVLPAPAPAPPPSPPPPGVQHTGLSPYTLFVASQKRRKREEDSSKSKKKKWIS 2591
Db 2923-----DNESVTTSEKSEAELEKNSSEKSAEKKSTS 2953
OY 2592-----TTSKETKKDT 2601
Db 2954 ADLEAADKAETEKSETGKET 2973

RESULT 13
A53577
Accession: slalogleycoprotein 1 - rat (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C:Accession: A53577
R:Mu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A:Title: Molecular cloning and sequencing of the muin subunit of a heterodimeric, bl
A:Reference number: A53577; MUID:94216302; PMID:8163496
A:Accession: A53577
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1630 <MUA>
A:Cross-references: GB:U06746
C:Keywords: glycoprotein

Query Match 2.8%; Score 402; DB 2; Length 1630;
Best Local Similarity 19.8%; Pred. No. 1.6e-07; Indels 362; Gaps 56;
Matches 296; Conservative 197; Mismatches 620;

OY 1205 NCEDRLPVKGTENGKK-----PSQOKLIERPVNKCSDIQLKNT-----TD 1247
Db 21 SCILLIPVNTSTTSARKTSTALPSSSTNPQMSQVSNPT--ASSYRMKNTQOASPMVTS 78
OY 1248 KKNNESESEKQORTSTQIINGKDNKPKIYLKGECLKETSRRVYSGVNEPKVANNINKI 1307
Db 79 SSITTLPOSHGTSMKTT-----RNPQTGTTEVTTLSASS--SDQOVETTSTQTL 129
OY 1308 IPEVDIKSLVKSASIRPFIINGDIVIMEDFERNSSFTKSHLSSPDAGNYSLETLPSS 1367
Db 130 SPDTTTSNAPRESSSP-----STVILTTTATBETSGDTGHTMAV 172

```


OY 1368 TRESSTQTPPSASCPESSVNOVEDMEIETSEV-----KVVTSPI---TSE 1413
 DB 173 TQOG-STPATTEISVTPSTQKMSPVSTFSTSTQOEITLILSQSHGGMKTRNPQTGTTE 231
 OY 1414 EESNLSNDPIDENGLPINKNEVNGESKRKYITEVTMTSTVALESKTVIVEGDKOT 1473
 DB 232 VTTTTLSSASSDH---PTSSPESTPGNTAPRTETSTTTTKVLMSTLQO--KLPGC--ST 284
 OY 1474 VVSSTENCA-----KSTVTTTTVTKLSTPGSGVDLISVGEOSKQTVTTTV 1522
 DB 265 LCTSTQOEITLTPQSOHTGIMKTTSRTQTTTPEVTTRITLSAS---SDHRQETSQITL 341
 OY 1533 TDSLTTTG-----GTLVTSMTVSKESYTRDKVCLKMFSRPRKTRSGTALPSYRKEV 1573
 DB 342 SPDTTTSNAPRESSPPSTSVILTHGREGTSGDGHMAVTQGSTPATTEIS-----V 396
 OY 1574 TKTSTKKSIFVLPNDLKLARKGGRVRYFNYNAKPALDIPYSPRPTFQITWRRLQ 1633
 DB 397 TPSTQKM-----SPVSTFSTS---TQ 414
 OY 1634 TVKSLAGVSLMLRLMLASLRMDMAKVPYGGGSTRT--ETSETETTTTETIKRDOVBY 1691
 DB 415 ETTTTL-----OSHTGGMKTRNPNORTTPTTEVTSTLSAS---SSD 453
 OY 1692 GIREFYCIARKIICPIGVPEPTPEPTPOKGLRSSALRKRPE-TPKOTGPYIETWVAE 1750
 DB 454 QVQVETTSRATL-----SPDTTTSNAPSVSSSPSTEGTSVDG--LTTAVTT 503
 OY 1751 EELEMEIRAFEREREKEKAAVEODAKRLRQKPYATITSTTSSTTSTISP---- 1806
 DB 504 DO-----STPATQGLTS--SQOTLVSPLSTS 531
 OY 1807 AQKVVAPISGSVTTGKMLVLTTKVGPATVTFQONKNHOFETATVWKGOSNGVVOQ 1866
 DB 532 TQETSTQETLTSQSHTGSMKTH--NPOTT--KTEVTTLTSLA-----SSDOVOYE 580
 OY 1867 QKVLGITPSSSTGTSQOFTSTFQPRAT-----VTIRPNTSGSGGTTNSNGVITGPDI 1918
 DB 581 -----TTSQTLSDATTTSHAPRESSPPSTSDILTTWASTEGTSGDGHMAVTQGS 634
 OY 1919 RFGMTVIR-TPLOQSTGLKALIRTPVMOFGAPQOVM-----QIRKO 1961
 DB 635 TPATTEISVTPSTQK-----MSVPSTFSTQETLTLISOSQHTGGMKTRNPNQTGT 687
 OY 1962 PVSIVASAPNT---VSTSTPGQSKLT-SATSTSNISASASQPRPOQGVKLTMAOLTQUT 2017
 DB 688 EYTTTLSSASSSQOVAETSSQTTLSPTDTTSHAPRESSPPSTSD-----MLTTTAST 741
 OY 2018 OGHGNGOGLTVVIGOGQOTGQLOI---IPQGVTVLPQPGQOLMO--AAMP-----NGTV 2067
 DB 742 EGTSGDGTGHTTAVTQGSIPATTOLTSTTFASQKMSVSTPTTSSIOELSTLHPQSHGTM 801
 OY 2068 QRFLEPTLATITTTASTTTTSTTAAGTGEOROSKLSQOMOVHODKTLPPAQSSVGA 2127
 DB 802 E-----TSSRPQTTSVSTSSSSPSGSTPVQTRSVTS-----SSDERTNPTSSGVSNTS 850
 OY 2128 KAOPTAPASAPPOPOFOPSPAPPOEVOT-----OPEVOTQTTVSHVPSAEP 2176
 DB 851 PATTEVLTPSTPESTPGNTAPRTETSTTTTKVLMSTLQOKLPTGSLGTSTPEVTT 910
 OY 2177 THAOSKPOVAAQOPQSNVQGSFVRVQSPQTRIRPST-----PSQLSPGQSOV 2228
 DB 911 TLSASSDOVQVETTSQTTLSPDATTTSHAPRESSPPSTSVILTTMASTESTSDTGH 970
 OY 2229 QTTTSG-PITPQPHSLQIPSGQPOQPOVQOSTQTQTLSSGOTLQOVVSSPSRPQ---- 2283
 DB 971 TAVTDQGSTPATTEISV-TPSTQKMSVSTVLTSTQELTSSQSTGKSGTSSKPOATTP 1029
 OY 2284 -----LQIOPO--POVAVPQLQOVQVLSQIOQOVAAQOIA 2319
 DB 1030 TEVTTSTLSSFRGSLFSARNCCLOATKRPPLAVVCLDPSSVPSLHMSSKPOATTPTEV 1089

OY 2320 QOSGVPQOIKILOPIQIOSSAVQTHQIQNVYVOAASVQEOLOQVQOLRDOQKKKKQO 2379
 DB 1090 TTS-----TLSFSRSGTQOTVSMETSSSGKITAPSTSSRRPVSATSDIFTTTDS 1141
 OY 2380 IEIKREHTLQASNOSEILIOKOVYMKHNAVIEHLKOKSKSTPAREBNQMIYCNQYMKI 2439
 DB 1142 TSGMNGHTLLTSHSVYTSR-----VASTTGLRSLTVAHROSTORSSTHSQ--SYL 1190
 OY 2440 LDKIDKEEQAAKKRRRESEVQKRSKONATKLSLALFHKRQOLRAELIKKRALDKDQ 2499
 DB 1191 TSMG-----ASSTSETSLTETATTEISLCLFTWTH-----CDRLD- 1226
 OY 2500 IEVQELKRDILKIKKEDLMQLOATVAAVPCPVYTPVLPAPPAPPPPPPGV 2554
 DB 1227 --LSMTSSSGLTFTKTDNDRSTALSATSLTLPAPs-TSTASRSSTVPPAPLPPOGI 1278
 RESULT 14
 S48478
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
 C:Accession: S48478; A26877; S27281; JC6123
 R:Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A:Reference number: S48478
 A:Accession: S48478
 A:Molecule type: DNA
 A:Residues: 1-1367 <ROW>
 A:Cross-references: GB:247047; EMBL:236061; NID:9603997; PID:9763364; GSPDB:GN00009;
 R:Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A>Title: Gene fusion is a possible mechanism underlying the evolution of SWA1.
 A:Reference number: A26877
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA5014.1; PID:g172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YAZ>
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA5015.1; PID:g172526
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Clares, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A>Title: Similar short elements in the 5' regions of the SWA2 and SWA genes from Sacc
 A:Reference number: S27281; MUID:89031230; PMID:3141213
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R:Landreth, M.G.; Bauer, F.F.; Marmur, J.; Pfeiffer, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A>Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseud
 A:Reference number: JC6123; MUID:96323237; PMID:8710886
 A:Accession: JC6123
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAAC94609.1; PID:g1304387
 C:Genetics:
 A:Gene: SGD:MUC1; STA2; MAL5; DEK2; SGD:S0001458
 A:Cross-references: MIPS:YIR019c; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:5-21/Domain: transmembrane #status predicted <TM>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>
 Query Match 2.8%; Score 401.5; DB 1; Length 1367;
 Best Local Similarity 21.2%; Pred. No. 1.4e-07;
 Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;

Qy	1135	DEBG--KIGCSE-----SNSILENSDPIVSIQDSBEDMIVQUNSINSISEQFT	1182
Db	191	DPGFGVWINDCNNOGCKSTSTTSSTSESSSTTSSESS-----TTSSTSESSST	243
Qy	1183	REODVAVLEPLKCELVGSESTGNC--EDRLPVKGTENGKPKPSQOAKLEBPVYNKESDOI	1240
Db	244	TSSSTSESSSTSTAPAPPTTSCKEKPPPTTSTCKEKKPP-----PHDTPCT	296
Qy	1241	KLKNITDKNNENRESEKKGORTSTFQJINGKDNKPKLYLGECLKELSESRYVSGVNEK	1300
Db	297	KKRTTTSKCTCKRTTTPVPTPSSSTPE---SSAPVPTPSSSTTSSSAPVTSSTPE--	350
Qy	1301	VNNINKTIPENDIKSLTVEKSIREFINGVDIMEDFERNSESEKSHLLSDAEGRND	1360
Db	351	--SSAPVP---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP	394
Qy	1361	SLETPSTKESDS--TQTPPSACPEPSNOVEDMELETSEVKVYSSPTSE----	1413
Db	395	VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV--TSSTTESSAPVTSSTTES	453
Qy	1414	-----EESN--LSNDPFDENGLPINKNEANVNGSKKRYTVEVTMTST--VA	1457
Db	454	SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV	513
Qy	1458	TESKTVLKEKGDKQTVVSTSTENCAKSTVTTTTTTLKSTPSTGSGVDITSKEOSKIV	1517
Db	514	TPSSSTTESSAPAPPPSSSTTESSAPVTSSTTESSAPVPTSS-----STTESSPT	568
Qy	1518	VTTTIVDLSLTGTLVYTMVFSKE--XSTRVKYKLMFSPPKRYRSGTALPYSRKFVVK	1575
Db	569	VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPAPPTPSSSTTESSAPVTS	628
Qy	1576	STKKSIFVLEPNDDLKLRKKGIREVYFVYNKAPALDIMPSPRPPTFGITWRYLQIV	1635
Db	629	STTESSA-----VPTPSSS-----TTE	647
Qy	1636	KSLAGVSLMLRLMALSIRMDMAKVPPOGSGSTRETSETELT-----TELIRK	1685
Db	648	SSSAPVPTP---SSSTTESSAPVPTPSSST--TESSAPVTSSTTESSAPVTSSTTE	701
Qy	1686	RDVQVPIREYCIKRIKIPICIVP---EPKPEPTPOKGLRSSAL-----	1728
Db	702	SSSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTESSS	761
Qy	1729	---RPRKPEPKOTGVYIETWVAEELIETAPAEAREKKAQAVEQOAKRELOQ	1784
Db	762	APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAVPTPSSSNTSS	821
Qy	1785	KP--TVATSTTS-----PTSSSTSTSPAKWVAPISGVSVTGKWLTTKVS--	1833
Db	822	APSTPSSSTTESSVVPVPTPSSSTESSS-----APVSSSTESSVAVPTPSSSN	874
Qy	1834	---PATVTFQOKNKHOTFATWAKOGOSNGSVQVOAKVLG-----IIPS-----	1875
Db	875	ITSSAPSSIPSSSTTESFST--GTTVPSSSKYPGSOETESVSTETTIVTKTTSVTT	933
Qy	1876	-----STGTSQ--OTFTSOPRATVYIRPNTSGSGGTSNSOVUTFGQIIRGM	1922
Db	934	PSSTTTTTTVCSTGNAGETSGCSPKVTYTTV--PTTTTSVTTSSTTTTIT-----T	985
Qy	1923	TVIRTPLD--QSTLG---KAIIRF--PVAVQCAPQVWQTIIROPISTAVS-----AP	1970
Db	986	TVCSGTGTSAGETSGCSPKTIITTVPCSTSPETASBESTTSPPTVTVVSTTVTTE	1045
Qy	1971	NVVSSTPOKSLGTSATGSNIOS--ASOPRPOQOV--KLMAOLITOLTOGHGNOGLT	2027
Db	1046	VSTSKRPGE--TITTEVYTKNIPITLYLTAPRPSVTVVNFPTTITTTVCSTGNNA--	1102
Qy	2028	VWIGOGGOTTGOLILPOQVTVL---PBGQOOLMAAMPNGSTVORFLFTPLATTATAS	2083
Db	1103	-----GFTSGGS--PKVTYTVPCSTGIGEVTTA-----TTLVTTA-----	1138
Qy	2084	TTTTTVSTTAAGTEBOROSKISPMOVHDKTLPRAOSSVGPAAKQOPOTAPSKRPPQ	2143

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:33:37 ; Search time 12.712 Seconds
(without alignments)
3426.253 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MVSEEEEDDDEDAETQDSE.....KIKGFKASRSHNNKLOSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100480 segs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1474	10.3	328	US-09-925-297-816	Sequence 816, App
2	982.5	6.9	238	US-09-729-835-80	Sequence 80, Appl
3	418	2.9	4019	US-09-738-973-425	Sequence 425, App
4	416.5	2.9	2665	US-09-864-761-34248	Sequence 34248, A
5	403	2.8	3256	US-09-919-172-98	Sequence 98, Appl
6	401.5	2.8	1367	US-09-801-368-108	Sequence 108, App
7	401	2.7	2344	US-09-815-242-12713	Sequence 12713, A
8	391	2.7	5179	US-09-922-217-1068	Sequence 1068, Ap
9	391	2.7	5179	US-09-833-263-1068	Sequence 1068, Ap
10	355.5	2.5	2478	US-09-815-242-5816	Sequence 5816, Ap
11	355.5	2.5	2478	US-09-815-242-12967	Sequence 12967, A
12	336	2.3	2843	US-08-681-219-32	Sequence 32, Appl
13	323	2.3	1596	US-09-902-432-4	Sequence 4, Appl
14	321.5	2.2	1400	US-09-764-176-7	Sequence 7, Appl
15	321	2.2	2063	US-09-735-367B-2	Sequence 2, Appl
16	320.5	2.2	1781	US-09-738-877-3	Sequence 3, Appl
17	318	2.2	2005	US-09-735-367B-3	Sequence 3, Appl
18	317	2.2	1363	US-10-124-557-52	Sequence 52, Appl
19	315	2.2	6281	US-09-815-242-12996	Sequence 12996, A

20	312.5	2.2	1140	US-10-124-557-104	Sequence 104, App
21	312.5	2.2	1404	US-10-124-557-2	Sequence 2, Appl
22	312.5	2.2	1404	US-10-124-557-62	Sequence 62, Appl
23	311.5	2.2	1049	US-10-124-557-58	Sequence 58, Appl
24	311.5	2.2	1313	US-10-124-557-142	Sequence 142, App
25	311.5	2.2	1354	US-10-124-557-48	Sequence 48, Appl
26	309.5	2.2	2368	US-09-815-242-5635	Sequence 5635, Ap
27	309.5	2.2	2368	US-09-815-242-12389	Sequence 12389, A
28	309	2.2	2441	US-10-109-886-8	Sequence 8, Appl
29	306	2.1	957	US-09-922-217-1065	Sequence 1065, Ap
30	306	2.1	957	US-09-833-263-1065	Sequence 1065, Ap
31	305	2.1	1320	US-10-124-557-46	Sequence 46, Appl
32	305	2.1	1320	US-10-124-557-60	Sequence 60, Appl
33	305	2.1	1361	US-10-124-557-40	Sequence 40, Appl
34	304	2.1	1314	US-10-124-557-50	Sequence 50, Appl
35	301	2.1	1038	US-10-124-557-74	Sequence 74, Appl
36	301	2.1	1270	US-10-124-557-44	Sequence 44, Appl
37	301	2.1	1311	US-10-124-557-42	Sequence 42, Appl
38	300	2.1	1022	US-10-124-557-84	Sequence 84, Appl
39	297.5	2.1	5795	US-09-815-242-12610	Sequence 12610, A
40	296	2.1	1537	US-09-801-368-104	Sequence 104, App
41	295	2.1	2442	US-10-109-886-10	Sequence 10, Appl
42	293.5	2.0	941	US-10-124-557-14	Sequence 14, Appl
43	289	2.0	1111	US-09-815-242-12955	Sequence 12955, A
44	288	2.0	1070	US-09-735-367B-6	Sequence 6, Appl
45	286	2.0	128	US-09-764-864-1571	Sequence 1571, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-816
Sequence 816 Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925, 297
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124, 270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (178)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (269)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-297-816

Query Match 10.3%; Score 1474; DB 10; Length 328;
Best Local Similarity 94.8%; Pred. No. 9.1e-58;
Matches 289; Conservative 0; Mismatches 14; Indels 2; Gaps 1;

QY 406 NKSFLAANEELIESIRAKGCDIDNFKSP--EETEKDKNETENDSDAENREFEQDQSL 463
DB 12 NVSLSGTNEELIESIRAKGCDIDNFKSPTEETEKDKNETENDSDAENREFEQDQSL 71
QY 464 EKSDSDKTPDDPEQSGSEVDFKSEKSNELSESPGAGSGASTRIITRLNPDPSKLS 523
DB 72 EKSDSDKTPDDPEQSGSEVDFKSEKSNELSESPGAGSGASTRIITRLNPDPSKLS 131
QY 524 QLSQVAAAHAHEANKLFEGKEVLYVNSGSEISRLSTKKEVIMKNNINNYFKLGQBGK 563
DB 132 QLSQVAAAHAHEANKLFEGKEVLYVNSGSEISRLSTKKEVIMKNNINNYFKLGQBGK 191
QY 584 RYVHNOYSTNSFALNKHQHEHDHDKRHLAHKFCPLPAGFEKMGNSVHSGKVLITSLRL 643
DB 192 RYVHNOYSTNSFALNKHQHEHDHDKRHLAHKFCPLPAGFEKMGNSVHSGKVLITSLRL 251
QY 644 TITOLENNITPSFLHPNMAHSHRANWIKAVOMCSKPREFALALAILCCAVKPVMLPTWRE 703
DB 252 TITOLENNITPSFLHPNMAHSHRANWIKAVOMCSKPREFALALAILCCAVKPVMLPTWRE 311
QY 704 FLGHT 708
DB 312 SLGHT 316

RESULT 2

US-09-729-835-80
; Sequence 80, Application US/09729835
; Patent No. US20010016647A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015P1
; CURRENT APPLICATION NUMBER: US/09/729, 835
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 09/257,179
; PRIOR FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: 60/056,270
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,271
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,247
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,073
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-729-835-80

Query Match 6.9%; Score 982.5; DB 10; Length 238;
Best Local Similarity 78.8%; Pred. No. 1.5e-36;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;

QY 253 PTLPAASQKRRREER-----DSSSKSKKKKMIISTYKSEKTK-----DT 2601
DB 3 PTLPAASQKRRREER-----DSSSKSKKKKMIISTYKSEKTK-----DT 2601
QY 2602 KLYCICKTPYDESKFYICGDCQNMVHGRVGIQSEALIDVEYVPCQGSTEDAMTVLT 2661
DB 59 ELVCICRTPYDESKFYICGDCQNMVHGRVGIQSEALIDVEYVPCQGSTEDAMTVLT 118
QY 2662 PLTEKDYGKLRVLSLQAHKMAFLEVPDNPADYGVYIKERMDLATMEERVQRRY 2721

DB 119 PLTEKDYGKLRVLSLQAHKMAFLEVPDNPADYGVYIKERMDLATMEERVQRRY 178

QY 2722 EKUTEFVADMTKIFDNCRRYNSDPSFYOCAEVLESEFPYOKLGFKASRSHNNKLOSTAS 2781
DB 179 EKUTEFVADMTKIFDNCRRYNSDPSFYOCAEVLESEFPYOKLGFKASRSHNNKLOSTAS 238

RESULT 3

US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738, 973
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match 2.9%; Score 418; DB 10; Length 4019;
Best Local Similarity 17.4%; Pred. No. 1.2e-10;
Matches 492; Conservative 320; Mismatches 995; Indels 1018; Gaps 103;

QY 254 CRVCHL-----GDLICETGSAVYHLECVKPRLEVEDEMOCVCYAHKVPVTCVARI 310
DB 65 CEACGATDPGRILLICDDCISYHYCLDPLQTVKGGKCMWCNHCATISAGLRC 124
QY 311 OKKKPYIRHPPIGYDSRRKKYFELNRLITEEDTENENKKIWIYSTKVQALIELDCLDK 370
DB 125 EMQNNITQCARPCA-SLSSCVCYRNR---EEDL-----ILQCRQC 161
QY 371 DYWEAELCTLLEMRREIRHRMDITBEDLTNKARGSKSFLAAN-----EELIESIR 422
DB 162 DRMHMAVCCQNL-NTEEEVENADIGFDCS---MCRPYMPSNVPSDCCESLVAQIV 215
QY 423 AKKGIDINVSPE-----TEKDKNETENDSDAENREFEQDQSL----- 471
DB 216 TKVKEIDLPKTYTQDGYCLTESGCTQLOSITVVPKRRKSKPKYLKTIINQNSVANVLOT 275
QY 472 PD-----DDEQSGSEVDFKSEKSNELSESPGAGSGASTRIITRLNPDPS 520
DB 276 PDQSEHSRDEMDSDREGLMDQDCKSE--SPEREAVDDETRGEGTVGKRRKRP-- 332
QY 521 KLSOLSKQVAAAHAHEANKLFEGKEVLYVNSGSEISRLSTKKEVIMK---GNINNYFKL 577
DB 333 -----YRPGIGGFVWRQSRFGQGTKRSVIRKDSGSGISQQLPC 372
QY 578 GQEGKYRVYHNOYSTNSFALNKHQHEHDHDKRHLAHKFCPLPAGFEKMGNSVHSGKVL 637
DB 373 RDDG----- 376
QY 638 ISTRLTITOLENNIPSSFLHPNMAHSHRANWIKAVOMCSKPREFALALAILCCAVKPV 697
DB 377 ----- 376

QY 698 LPIWREFLGHTRLHMTSLEREKEKVKKKKEOE--RET-----MOQATWVYTPPVKH 751
Db 377 ---WSQOLDPTLVDESIVS--TESTEKIKRYKRKKKLEETEPAYLOEAFEGDILLDTSR 432
QY 752 QWVKOGEEYRVYGYGWMGSMISTKTHYRFVPLPGNTNVTNYRKSLEBGTKNMNMENDESD 811
Db 433 Q-----SKISIDNLSDEGAQLTYKTNNM-----TGFIDPSID--P 465
QY 812 KKKCSRPKIKIIEPDSEKDEYKSDAAGADONEMDISKITEKQDQYKELLSDSDRP 871
Db 466 LSSSSAPRK-----SCTHGPADPLADISEVL--NTDDILIGITSDLSKS 510
QY 872 CKEEPR--EVDDDMKTSHVNCQESSQYDVYVNSEGHMTSTYKTKKSKSLDGLERRIK 930
Db 511 VHSDDGPVTDPPSLPQPNVNOSS-----RPLSEBOLDGIL----- 547
QY 931 OFTEKORLEKIKLEGKIGK--IGKTSTNSKNLSPEVITKAFEGQSDSMQOEOP-- 987
Db 548 -----SPLDKMYTDGALLGKIKYKIPELGKDVEDLFTAVLSPANTQPTPLPQPPPTQ 601
QY 988 -----NANNDOPEDLIQCGSQSDSVLRMSDP-----SHTNKLYPEKRVLDVSI 1034
Db 602 LPLIHQDAFSRMLNGLIGSSPHLPNLSLPGSGLTGTFSAIQSSYPDAR--DKNSAF 659
QY 1035 SPETKCPKXN-----SIENDIEKVSDLASRQOEPKSKTKGNDEFIDDSKIASADDIG 1088
Db 660 NPASADPNNSWTSASPTVEGE-----NDTMSNAQRSTLKWEK-----EALAGENATVA 707
QY 1089 TLICKN--KKPLIOEESDITVSSSKSALHSSVPKSTNDRDATPLSRAMDEGKIGCD--SE 1145
Db 708 PLYLTINPINKLEFPDMTKRYKO--IAKLMRKASSQERAPYOKARDNALRIKVVQ 766
QY 1146 SNTSLENSDDTVSIODSE-----EDMIYONSNEISSEOFTRREO-----DY 1187
Db 767 SMDSMKROQOOSIDPSSRIDSELFKDPKQRESEH--EQEMKFRQOMQKSKQAKIETAT 825
QY 1188 EYLEPJKCE-----LVSGESTGNCEDRLVYKTEANGK--KPSQO----- 1225
Db 826 OKLEOVKNQOQOQOQOFGSOHLLVQSGSDTPSSGIOSPLTPPOGNGNMSPAOSFKELEF 885
QY 1226 -KLEERPYNKCSDOIKLKNTT-----DKKNENNESKKGKQRTSPQING 1271
Db 886 TKQPESTPTSTSSDDVFAVQAPPPAPASRLPIDLSLSOAOQTSOPPSPOVSSPGSSNR 945
QY 1272 DNKP--KIYKL-----GECIKIESESRYVSGVNEPKVNNINKI 1307
Db 946 PSPMPDYAKMGTPRPPVGHFSFRNSAAPENCTPLSSVSRPLQMN-----ETANRP 1001
QY 1308 IPEBNDI--KSLVYKESAIRPFTINGDIAMEDFENRNSSEKSHLSSDAEG----- 1356
Db 1002 SPVRDLCSSSTTNNNDYAKRPDTPRPVMTDOEPKSLGLSRSPVYSQGTAKGPIAAGTSDH 1061
QY 1357 -----NYRDSLETLPTSKESDSIOTTTSPASCPES-----NSVNOV--E 1393
Db 1062 FRKPSRADVFORQRIPDSTARPLLT--PAVLDSGPEPKTPMQPPSSODPYGSVQASR 1120
QY 1394 DHEIETSEVKYTSPISTIEEESNLNDFIDENGLPIKNKENVNGESKRTVTEVTMT 1453
Db 1121 RLSVDYERPALTRPIDNFHSH--NOSNDYQOP--PLTRHPAAN-----EFAFAP 1167
QY 1454 STVATESTVIVVEKQDKOTVVSST-----ENCAKSTVTTTTVTYKLTSPSTGSGVDII 1508
Db 1168 SFAFSOPGTSISRPTSDPPSOPPGTPRPVAVDSQSSGTARSNTDYSOPPGPRPTTYVD 1227
QY 1509 SVKEOSKTVVTTVTDLSLTGTGTLVTSMTYSKEYSTRKVKLMKRSRKKKTSGTALS 1568
Db 1228 PYSQOQPTPRPSTQDLEFYTP-----VTNQRHSDPVIAHPGTPRPGISVSPSOPAPPRPR 1283
QY 1569 YRKFTYKSTKSIIFVLPNDLKLAKKGIRVYFYN--YNAKPALDIMPYPSRPTFGI 1626
Db 1284 ISEGFRRSMTFRVLMRPND-----PFLQANQNRPAL--PGPLVPR----- 1323
QY 1627 TWRYLQTVKSLAGVSLMLRLMLASLRMDMAKAVPPGGSGSTRTETSETEITTEIIRK 1686

Db 1324 ----- 1323
QY 1687 DVGPRGIREYCIKIKIICPIGVPEPKETPTPOKRLARS--SALARKRETPKOTGPVLI 1744
Db 1324 -----PDTCSQTPRRPGPLSDTSTFSKVSFSAARDPJDOSPMPR 1361
QY 1745 ETWVAEELELMEIRAFABREVEKEKAQAVEQAKKRLBQKP--TVIATSTTSPSSPT 1801
Db 1362 RS-----QSDSEFGTSQTAHDVADQPRPFSESGSCFCASSNSPMHSOG 1401
QY 1802 STTSPAKVMVAPISGSVTTGTGMVLTITKVGSPATYTPQONKPFHQTFATWVQOGSNG 1861
Db 1402 QQFSVGSQ--LPGVPPTSGYTD-----QNTVMAADATEKLRQOKLRE 1444
QY 1862 VW-----QVOOKVLGILPSSGTGS-----QOFT----- 1885
Db 1445 IILQOQOQKKIAGROKSGSDSPAVHPRGLQHWQOENNOATRRPPRYPGNIRSPVAP 1504
QY 1886 -----SFQPTATVTIRPN-----TSGSGGTTSSNS-----QVITG-- 1915
Db 1505 PLGPRYAVFPKQDRCGYPPDVASMGMRPHGFRPGSGSHGTWPQOERFLVPPQIOGSG 1564
QY 1916 --PQIRPGMTV-----IKTPL-----QOSTLICKALI-- 1939
Db 1565 VSPQLRRSVSDMPRLNNSQMNPNVGLPQHFSPOSILPVQOHNILQAVYTELRHAPDGR 1624
QY 1940 -RTPVNVVQGA-----POQVMTQI----- 1957
Db 1625 QRLPESAPGSGVVEASNLRHGNFIRPDEPRGRHNDPMKRRPQGLPNOPLVHPDLEQVP 1684
QY 1958 ---IRGQPVSTAVSAPNTV-----SSTPGQKSLTATSTSNIOSSA----- 1995
Db 1685 PSQOEGHSHVSSSMVMTLNLHPLGGEFSEAPLSTSVPSBETSDNLQITQPSDGLEKL 1744
QY 1996 -SQPPRQOGQVK-----LHMAQLTOLTQGHGNGGLTVYI 2030
Db 1745 DSDDPVKELDVLDGVEYKLDLDEDLNLTDEGKVELDITLNLNETNPNLIDLL 1804
QY 2031 QGQGGTQGLQLOLIP-----QGVTVLPGPQOLMQAA 2061
Db 1805 R-----SGERDIIATYDPELMDGDKSMNEELDLPIDKLDNQCVSEPKKKEQ----- 1854
QY 2062 MPNGTVQRFLEPPLATATTAATTTTSTVSTTAAGTEBQROSKL-----SPQM 2108
Db 1855 -----ENKTLVLSDKHSPOKSKVTNEVKTEVLSPMSKYESKCEKTEKNDKNDVDTPCS 1909
QY 2109 QYHOKTLPPAOSSVGPA-----KAOPQTAQPSARPOQTQPOSPAPQREVQOTQPEVQ 2161
Db 1910 QASAHSDLNDGEXTSLHPCDPDLFEKRTNRETGAPSAN--VIOASTQDLPADQVYNSCGITG 1968
QY 2162 TQTTVSHVPEAQPT--HAQSSKPOVAAQSOPOSNOGQSPVVRVQSPQTRIRPSTPS 2218
Db 1969 STPVLSLLANESKNSDILRPSGSPPTPLPASFSNHVSSLPPF-----IAP----- 2015
QY 2219 QLSPGQ-----QSOYQTTTSQPTPIQPHNTSLAQIPSQ-----GQPOSQ-- 2255
Db 2016 ---PGAVLNDAMANSNTVTVSRVNHVFSQGVNPRGL---IPGSGTVNHLISGCTKRPATQNG 2069
QY 2256 PQY--QSTQTLSSGQTL--NOVSVSSPSKRPOLQIQPORQVAYVQLOQOQVULSQIOS 2311
Db 2070 PQTSQSGTSMGSPQOOLMIPQTLAQOQNRERPLLEQDPLILDILQEROQOQOQOMQDA 2129
QY 2312 QY-----VAQIOAQOOS--GVPOQIKLOLPIQIOQSS 2340
Db 2130 MKRQSEPPFPNIDPDATIDPILKAKMAVALKINKVMAQNNLGMMPVMSRPFPGQVYV 2189
QY 2341 AVQTHQIQWY--VTVQOASVQEOLOQVQO-----LRPOQOQKKKQOQOIEIKREHTLOA 2390
Db 2190 GTQNSQGNLNGPRAIPQDSISITHQISRPNPNPGPFVNDQSKQOYEBMLQ----- 2240
QY 2391 SNOSEIIQOQVWKNHNAVYIEHLKQKKSMTPAERENQRMIVCQVYKVIILDKIDKEEKA 2450

Db 2241 -ETQOOLQMOQKYLEQIGAHRSKRALSAKOR-----T 2273
QY 2451 AKRRKEEVEOKRSKONATKLSALLFKHKEOLRAELIKKRALLDKQLEJOELKRD 2510
Db 2274 AKKAGREFPEEDBAFQKHTYEQOSWQKOLEQIRK-----QCKEHAELIEDY 2320
QY 2511 KIKKEDLMQLOATAVAAD-CPVTPVLPAPAPPPPPPPGV-----QHTGLLST 2562
Db 2321 RIKQOQ---QCAMAPPMPSVQOPPLIGCAPPTFMSCPTFFMVPQOLQOQHTVVISG 2377
QY 2563 PTLPLV 2567
Db 2378 HTSPV 2382

RESULT 4
US-09-864-761-34248
Sequence 34248, Application US/09864761
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-x-1
CURRENT APPLICATION NUMBER: US/09/864, 761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180, 312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207, 456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632, 366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263, 6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236, 359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234, 687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608, 408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774, 203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34248
LENGTH: 2665
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL034555.2

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
OTHER INFORMATION: EST_HUMAN HIT: A0117052.1, EVALU0 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P08640, EVALU0 3.00e-10
US-09-864-761-34248

Query Match 2.9%; Score 416.5; DA 10; Length 2665;
Best Local Similarity 18.5%; Pred. No. 8.9e-11;
Matches 544; Conservative 422; Mismatches 1145; Indels 825; Gaps 132;

QY 7 EEEEDGAETQDSEDE-----EDEMEDDDSDYPERMEDDDDAASYCTESSRSHS 59
Db 170 KNEKTKERTFDEPERVERERRLIRKEVEKTKTDQ----- 205
QY 60 TYSTGRRKPRVHRPRSPILE-----EKDIPLEPKS-----SEDLMPVNEHI 104
Db 206 -----KRGKGVHSPSSOSSSETDOENEROSP-EKPRSCNKLRSRKADEKGIATKRNLEL 257
QY 105 MNVAIYEVLRNFGVLRSLSPREFEDCAALVSOECTTMAEHNVLKALVLEEDTSNT 164
Db 258 MFCVVLTRVKEKEGKVIDHTV-----EKIKAKLNDITYKSSALDQKLOYQOT 305
QY 165 TFGPADLK--DSVNSTLYFLDM--TWPEVLRVYCESD--KEYHNVLRYQAEADYRGY 218
Db 306 EPAKSDLSKLESYRMKVPKESGLSHVEYV-----EKEGRLAKRKHILKPEQPAD----GVS 357
QY 219 ENKIKVLFVDPFLTNTARELMSE---GVLYQDHCVCYKGLDGLCCFCGSAYVHL 275
Db 358 AVDLKLEAKRRFRADSNLAEKQKPEVKKSSPEMD-ARVLSKKOPDVSREVIILRIG 416
QY 276 EGVKPPV--EEVEDEMOCEVCVAHKVPGVTQVAETIOKKRPYIRHEPIGYDGRSRKYWF 333
Db 417 EAERKVPKRIILKRESKIKLIDLNTVYASPKD-QELAS-----ISVSGSGSPSSD 466
QY 334 LNRLL--IIEEDTENENEKKIWIYSTKVOALIELIDCLDKDYAEELKILIEEM--REET 388
Db 467 LQARLELAGESEVNOEVOGSKRPISKPOLQO-----QYLDQGGPREBY 512
QY 389 HR-HMDITEDLTAKAGSNKSFLLAANEELIESIRAKG-DIDNVS--PEETEDKME 443
Db 513 RKNVCSLRDETPEKSGQEKSHSVNTEKI-----GIDIDHTQSYRKQMSRRKQ 564
QY 444 TENDSKDAE-----KNREFEEDOST-----EKSDDKTPDDDEQOK-SEVGDF----- 486
Db 565 MEMEIKASEKFGSPKDVDEYERKSLVHEVGKPPQVDYDTSPPSKKKRMDHVDFOICTKR 624
QY 487 -KSEKNGELSE-----SPGAGKA-----SGSTRITRLRNP--DSKLSQKSOQ 529
Db 625 ERNVRSSRQISEDSERTGSPSVRHGSFHEDEDPISPRLLSYKSGPKYDEKVLPSNIT 684
QY 530 YAAAHAENAKLFKEGKVL-----VNSGELISRLST-----K 562
Db 685 VRESESLKFNPDYSSRREOMADMAKIKLSVLNSENDELNRMDQKODAGRGDVSFPNSIR 744
QY 563 KEVIMKGNINNY-----FLGQEGKYRVYHNGYSTNSFLNKH-----QHRD----- 605
Db 745 RDSLRKRSVRDLEPEGVPSDSDEGE-----HKSHSRASALYESSLSLTLRREKRLAE 800
QY 606 HDRK-----RHIAHKFCL--TPAGEFKWNSGVSGVLTISTRLTITQLENNIPSS 655
Db 801 RDERLSSLSERNKFYSFALDKTTP-----DTKALLERAKSSLSSSREN--WS 846
QY 656 FLHPNASHIRANKIKAVQMSKPRERFALLALECAVKKPVVMLPIYREFLGHTRLHRMYS 715
Db 847 FL--DWDSEFANFRN-----NKDKE-----KVDASAPRPI--PSW--YMKKKKIR--- 884

GENERAL INFORMATION:
APPLICANT: Faris, Mary
APPLICANT: Turner, Christopher M.
TITLE OF INVENTION: PROSTATE CANCER MARKERS
FILE REFERENCE: PA-0036 US
CURRENT APPLICATION NUMBER: US/09/919,112
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/222,469
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PERL Program
SEQ ID NO 98
LENGTH: 3256
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incycle ID No. US20020119463A1 2700132CD1
US-09-919-172-98

Query Match 2.88; Score 403; DB 10; Length 3256;
Best Local Similarity 18.5%; Pred. No. 4.3e-10;
Matches 523; Conservative 371; Mismatches 1011; Indels 916; Gaps 130;

QY 339 IIEEDTENENKIKIMYSTKVQALJIDCLDK---DYWEALCKLIEEMREETHIRMDI 394
DB 432 VLPETEIHNEPFLTLMTLOVERKIOKDSLSKPEKLGTTAGOMSGSLPGLSS-----VDI 486
QY 395 TE--DITNKARG-----SNKSFLAANEILE-----SIRAKKDIDN-----VKSPEE 436
DB 487 NNGSGINESGIFLKRARRVFGGHLRPELFDENLPPNTPPLKRGAEAPTRKRSIWMHPPV 546
QY 437 TEKDKNEFENSKAEKNRE---EFEQSL---EKSDDKTP---DDPEQCKSVGDFKS 488
DB 547 LKTIKEPOPSPGKQESSEIHEVKAOSLVISPPAPSPRTVPAVSQORRSCCTAASS 606
QY 489 EKSNGELSESPGAGGASGSTRITRLNPDPS-----KLSQLKSOQVAAAHBANKLFK 542
DB 607 SKSQTEVFK-----RGERVATCLOKRVISRSQHDILQMICSKRSGASBAN-----654
QY 543 EGKEVLVNVSGEISRLSTK---EVIMKG---NINNTFKIGQCKRYV---YHQYQSTN- 593
DB 655 ---LIVAKSWADYKILGAKOTQTVIKHGPORSNNKRRPAPPKPVGEVHSGFSTGH 710
QY 594 ---SFLNKHQHEHEDDKRR---HLAKHFLTPAGEFK-----WNGSVGSKYLIT 637
DB 711 ANSPCTITIGRAHTEKVVAPAPRYVNLNFI SNOKMDFKEDLSGIAEMFKIPVQEQPOLT 770
QY 638 ISTRLLITTOLENNIPSSF-----LHPNMAHSHRAMIKAVQOMCSK--PREFALALAI 687
DB 771 -STCHIAISNSENILGQFOGDSGEPLPTSESFGSNVFFSAQNAKQPSD-----822
QY 688 LECANVPVVMPLTWRFEIIGH---RLHRTMSIRE-----719
DB 823 -KCSASP---PLRQOCIRENGNNAKTPRNTYKMTSLTKISDTEPESKTVSVNRSGR 877
QY 720 --EKKEVKK--KEKKOEEETK-----QOATWVYTPPVKHQVKKQGEERYVGY 766
DB 878 STEFNICKLPVESSEETNEIIVCIIKRGOKATLLOO-----RRGEMKEI---925
QY 767 GGSWISKTHYRVFKPLGNTNVYKRSLSBGTKNMDENDENDESK-----RKCS--RSP 819
DB 926 -----ERPFET---YKENIELKEN--DEKKAMKRSRTWQKCAPMSDL 964
QY 820 KIKIEPPOSE--KDEVKG-----SDAAKGADONEMDISKITEKKDOQVKELLDSDSKP 871
DB 965 TDLKSLPTELKKTARQONLLOTODHAKAPKSEKGLTKM-----P 1006
QY 872 CKE-EPMEVDMDKTESHNCOESSQADVNVNSE-----GPHLRTS-----911
DB 1007 CQSLDP-----EPINTPTHTKQOLKASISGKVEELAVGKTRTSGETHHTHREPAGDG 1062
QY 912 -----YKKKTK-----SSKLGLLEERRIKOPTLEEKORLEKIKLGGIKIGTGTSTNSK 961

DB 1063 KSLRPFKESPQIILDPAAAVTGM--KKMPTREKEAQSLEDL---AGEKELFQTPGSPSE 1117
QY 962 NISESEVITKAKEGQSDSMRQDSNANNDQPE--DLIOGQSQSDSVLRMSDPSTNNK 1020
DB 1118 SMIDEKT---TKIACKSPPESEVDPTSTKQWPKSLRADVEEELAKRLTTPSAGRAM 1174
QY 1021 LVPKDRVLDDVSI-----RSPETCKPQNSIEN-----DIE 1051
DB 1175 LTPKPAQGBEKDIAKMGTPVQKLDAGTLPGSKROLQPKRKAQALBDAFKELFQTP 1234
QY 1052 EKVSDDLASRGQ-----EPYKSKTGNDFTIDSKLASADITGLI--CKNK 1095
DB 1235 GHTEELVAAGKTTKIPCDSPQSDPVDTPSTKOR-----PKRSIRKADVEGELLACRNL 1288
QY 1096 KPLIOESDPTVSSKSAIHSVPKSTNDRD-----APPLSRAMPEGKLGDSSENSYL 1150
DB 1289 MP-----SAKAMHTPKPSVGEKDDIIEVGTIVQKL-----DLT 1323
QY 1151 ENSSDTVSIQDSSEEDMIYQNSNESISEQFRPREQDVEVLEPLKCELVSGESTG--NCEDR 1209
DB 1324 ENITGSKRRRQPTKEBAQALBLEDLTFKELFQTPGHTEEA-----VAAGKTKKMPCESS 1376
QY 1210 LPVKGTEANGKPSQOKLBERPVNKCSDQIKLKNITDKKNENRSEKKGQRTSTFOQN 1269
DB 1377 PPSADPTPTSTRQPKPTPEKRDVQKELSAK-----KLTQTSGETTHTDKVP 1424
QY 1270 GKNNKPRILYKGLCEKLEISRVSGNV--EPKVNINKIIPENDISILVKEAARFYN 1328
DB 1425 GGEKCSINARETAKOKLDAASVYTSKRIKPKK--EKQAPLEDLAWG--KELFQTP---1477
QY 1329 GDVIMEDFENRNSSETKSHLLSSDAEGNYRDSLETLPTSKESDSQTOTTPSASCSENS 1388
DB 1478 ---VCTDKPTTHEKTYKIACRSQPD-----PVDTPSSKPSQSKR 1513
QY 1389 VNOVEDMEITSEVKKTYTS-----PTSEESNLNDFTIDENGLPINK--NEWV 1436
DB 1514 SLRWVDEEEFALRKRTPKAGRAMHTPKPAVGEKIYYA--FM---GTPVQKLDITENL 1568
QY 1437 NGESKRTVITEVYTMSTVA-----TESKTVIKVEGDKQTVSVSTENCAS 1484
DB 1569 TGSRRRLQPTKEKAQALBDAFKELFQTPGHTSEMT---NDKTAACKSSQOLDLKN 1625
QY 1485 TVTTTTTVTKLSTPSTGSGVDIISVKEQSKYVTTVTDLSLTGTGLTVTSMTVEKVS 1544
DB 1626 PASSKRLKTSIG--KGVGVEELIAGKLTQTSGETHTHEPTGQSKSAKAMESPKOI 1683
QY 1545 TRDKVKLMKTSRPKKTRSGTA-----LPSYRK--FVTKSTKSIFFVLPNDD 1588
DB 1684 LDSAASLTGSKROLRTKKGSEVPEDLAGFLFQTPSHRKESMTNKTTKVYSRASQPD 1743
QY 1589 L-----KTLARKGIREVPYNNYNAKPAIDIMPPSPRPFGITWRYRLQTVKS 1637
DB 1744 LVDTPTSSKQPKRSLKADTEE-----EFLARQQTSPSAG---RAMHTPKP 1787
QY 1638 LAGVSLMIR--LIMASLKMDDMAKAVPPGGSTRTETSETETITTEILIKRDOVPGYR--1694
DB 1788 AVGEKDNITFLGPPVKLOPQNL--GSNRRLOQTREKKAQALEILT-----GREL 1838
QY 1695 -----FEYCIRKTIIC--PIGVP--ETPKETPPQKGLRSS-----ALRPKRP 1733
DB 1839 FQTECTDNPTTDEKTTKILCKSPQSDPADIPPTTKRPRRSLLKADVEEFLAFKRLTP 1898
QY 1734 -----ETPKQTPGVIIETVAEELELMEIRAP---ABRVE-----KE 1768
DB 1899 SAKAMHTPKRAA-----VGEER-----DINTFVGTPVVKLDLGLNLPKSKRRPQTPKE 1946
QY 1769 KAAQAVEQAAKRLDQOKPYVATATSTSPSTSTSTSPKAVVAPVSGSVTTGKAVLT 1828
DB 1947 KAKALEDLAGEKELFQTPGHTSEMTDKITIEVSCSKSPQDPVATP-----ISSKRLK 2000
QY 1829 TKVSPATVTPQOKNKFHQEPATWVKQGSNSGVVQVQOKVL--GIIPSSGTSGQOTFTS 1886

```
Db 2001 ISLGR-----VGKKEEVLPGKLTQTSKTTQT--- 2028
Qy 1887 FOPRATVTRIPNTPSSGGT-----TSNSQVI-----TGPOIRPGMTVTRP-----L 1929
Db 2029 -----HRETADGGSIAKFAKESAKOMLDPANVTGGMEMW-----RTKEKAOSL 2073
Qy 1930 QOSTLAKAIRPVPVWOPAPQOVWTOIIRGQPVSTAVSAPNTVSTPQOKSLTSTSTIS 1989
Db 2074 EDLAGFKELFQTPDHTEESTDDTKTKI-----ACKSPPESSMDTPTST- 2117
Qy 1990 NIOSSASQPPRRQOQGVKLT--MAOLTOITQHGNGGLTVVYQSGOGTTGQOLLP--- 2044
Db 2118 -----RRRPKTPPLGKRDIIVEELSLKOLTLQ-----TTHTDKVPGE 2153
Qy 2045 -OGVTVLPQGOOLM-----QAAMPNGTVQ-----RFLF--TPLATVATTAS 2083
Db 2154 DGINVFRETAOKLDPASVYSGSKRQPRTPKQKAPLEDLAKELFOTPTCTDKPTTH 2213
Qy 2084 TTTTIVSTTA-----AGTGE--QROSKLS--POMOVHODKTLPPQSSSVGAPAKAPOTA- 2134
Db 2214 EKTTKIACSPQPDVPYGTPTIEFPOSKRSLRKADVEEESLALRKRPVSGKAMDTPKPAQ 2273
Qy 2135 -----QPSARPQOTQPOSPAPQEVQOTQPEVQOTVYSS-----HW 2171
Db 2274 GBEKDKAMGTPVOKLDPGLPGSKRWPQTPKE-----KAQLEDLAGFKELFQTP 2326
Qy 2172 SEAPQTHAASSKPOVAASQPOSNVQSGSPVRSQSPQTRIRPS----- 2215
Db 2327 GNDKPT--TDEKTTKILACKS--PQ-----PDYDTPASTKQKRRKRLKADVEEFLALR 2377
Qy 2216 --TPS-----QLSPGQSOQVOTTTSQPIPIQPHISL--QIP--SQGQPOGQOP----- 2256
Db 2378 KRTPSAGKAMDPKPAVSEKNNINTEFETPVOKLIDLGNLPGSKRQPRQPRKEKALEDL 2437
Qy 2257 -----QVQSSQNTLSSQOTLQNVSSPSRQOLIQQRPQVYVAP-----QLQOQVY 2305
Db 2438 VGFKELFQPRGHTEESMTDKITTEVSCSPQPSFETSSSKORLKIPLVKKMKEEPLA 2497
Qy 2306 LSOI--QSOVNAQIQAOQSGVPOQIKL--QLPIQIOSSAVQTHQIONVTVQAASVQEQ 2361
Db 2498 VSKLTPTSETTQTHTEPFGDSKSIKAFKESPQOILDPAA-----SVTGSRRQLRTR 2549
Qy 2362 LORVOQLRQOQKKOQOIEIKREHTLQASNOSEI-----IO 2398
Db 2550 KKRARLEDLVDFKELFSAPGHTEESMTIDKNTKIPCKSPPELTPTATSTKRCPKTRLR 2609
Qy 2399 KOVVMHNAVIEHLKOKKSMT-----PAERENQMIYCNQVMKTIILDKIDKEQOAK 2452
Db 2610 KEVKEELSAV--ERLTQTSQOSTHTEKPEASGDEGIVL-----KORAK 2651
Qy 2453 KKRREESVQKRSKONA-----TKLSALLFKHKKEQLRA----- 2485
Db 2652 KKPNPVEEPPRRRRPAPREKAPLEDLAGFTLESTSGHTQESLTAAGATKIPCESPPL 2711
Qy 2486 ELIKKRALLDKLOLEVOBELKRLKIKKEKDLMLQAO----- 2523
Db 2712 EYVDVTAASKRHLRTRVOK-----VOVKEEPSAVKFTQTSGETTDADKEPAGEDGIKAL 2766
Qy 2524 ---ATAVAAPCPVPVPLVAPAPAPRPP-----PPRGVQHTGLSTPPLVPAVSQR 2572
Db 2767 KKSAPQTPAPASVYGSRRRRPAPRESAOAIEDLAGFKDPAAGHTEESMTDOKTKTIPCK 2826
Qy 2573 KREEEKSSSKSKK-----KMTSTSKETKKPYKLYCICKTGYDES 2614
Db 2827 SSPELEDTATSSKRRRRTRAQVKEVELLAVGKLTQTSGETTHND-----KEFVGEQ 2879
Qy 2615 K 2615
Db 2880 K 2880
```

RESULT 6
US-09-801-368-108

```
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: prt
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108
```

```
Best local Match 2.8%; Score 401.5; DB 10; Length 1367;
Best local Similarity 21.2%; Pred. No. 1.9e-10;
Matches 289; Conservative 108; Mismatches 564; Indels 339; Gaps 56;

Qy 1135 DFEQ---KICDSE-----SNSTLENSDPTVSIDQSSBEDMIVQNSIESIQFRT 1182
Db 191 DFGFVWINIDCNNCGTKSTTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST 243
Qy 1183 REDDVEYLEKCELVSGESTGNC--EDRLPVKGTENGKPKSQOKLEERPYNKKSDDI 1240
Db 244 TSSTSSSTSSSTTAPATPTTCTCKEKPPTTCTCKEKKPP-----PHHDTPTCT 296
Qy 1241 KLTNTDKKNRESEKQKQRTFOINGKDNKPKYILGCEKLETSESRAVSGANERP 1300
Db 297 KKTITTSKTCRTKTTTPVPSSSTTB-----SSAPVPPSSSTTBSSAPVSSSTTB 350
Qy 1301 VNNINKIIPENDIKSLVKSARPTINGDVIMEDFNERNSSSTKSHLSSDAEGNYRD 1360
Db 351 --SSAPVP--TPSSSTTBSSAPVTS-----STTBSSAPVTSSTTBSSAP 394
Qy 1361 SLETTSTKESDS---TQTTTPASCPESNSVQVEDMELETSEVKVYSSPTSE---- 1413
Db 395 VPPSSSTTBSSAPVTSSTTBSSAPVTSSTTBSSAPVTSSTTBSSAPVTSSTTBSS 453
Qy 1414 -----EESN--LSNDFIDENGPIPKKNENVNGESRKTYITEVTMTSP--VA 1457
Db 454 SSAPVPPTSSSTTBSSAPVTSSTTBSSAPVPTPSSSTTBSSAPVTSSTTBSSAPVP 513
Qy 1458 TESKTIKVEKQKQTVVSTENCASKSTVTTTPTVTKLSTPSTGSGVDIISVKEOSKTV 1517
Db 514 TPSSSTTBSSAPAPPTSSSTTBSSAPVTSSTTBSSAPVPTPSS--STTBSSSTP 568
Qy 1518 VTTTVVDSLTGTLVTSMTVSK--YSTRDVKYLMKRSRPKTKTSGLALPBYRKFVYK 1575
Db 569 VTSSTTBSSAPVPTPSSSTTBSSAPVPTPSSSTTBSSAPAPPTPSSSTTBSSAPVTS 628
Qy 1576 STKKSIFVLVNDLKLKARKGIREVPYFNYNKAPALDIMPYSPPTFGITRYRLQYV 1635
Db 629 STTBSSA-----PVTPSSS-----TTE 647
Qy 1636 KSLAGVSLMLRLMLASLRMDMAKAVPPGGGSTRTESETETTT-----TEIIRK 1685
```

```
Db 648 SSSAPVTP-----SSSTSSSAPVTPSSST-TESSSAPVTSSTESSAPVTSSTTE 701
Qy 1686 RDVGPYIREYCIKRLICIGVP-----ETKEETPPORKGLRSSAL----- 1728
Db 702 SSSAPVTPSSSTESSSAPVTPSSSTESSSAPVTPSSSTESSSAPVTSSTESS 761
Qy 1729 -----RKRREPMPQGTGVIIETWAELELMEIRAFAEVEKEKAQVDAQKRLLEQ 1784
Db 762 APVTPSSSTESSSAPVTPSSSTESSSAPVTPSSSTESSSAPVTPSSSTESS 821
Qy 1785 KP--TVIATSTTS-----PTSSTTISPAOKVMVAPISGSVTTGKMLVTTKVS-- 1833
Db 822 APSSTPSSSTESSSVDPVPPSSSTESS-----APVSSSTESSVAPVPPSSSSN 874
Qy 1834 -----PATVTFQOKNKHOFATVWKGQNSGVVQYQKVLG-----IIPS----- 1875
Db 875 ITSSAPSSIPSSSTESSST-GITVTPSSSKYPGSGTETSSTETTIVPTKTTTSVT 933
Qy 1876 -----STGTSQ--QTFSTFOPRTATVTRPMTSGSGGTTNSOYITGPOIRPM 1922
Db 934 PSTTTTTCSTGTSAGETTSAGETTSAGETTSAGETTSAGETTSAGETTSAGET 985
Qy 1923 TVIRTPLO--QSTLG--KAIIRT-PVMYQPGAPQOVMQIIRGQVSTAVS-----AP 1970
Db 986 TVCSTGTSAGETTSAGETTSAGETTSAGETTSAGETTSAGETTSAGETTSAGET 1045
Qy 1971 NTVSTGQOKLISATSTSNIOST--ASQPPRPOGOV-KITMAQLQOLQOHHGNGOLT 2027
Db 1046 YSTSTKPGCE-ITPTFVTKNIPTYLLTIAPTPSVTVTNPTPTTTTTCSTGTSNSA-- 1102
Qy 2028 VVIGOGGOTGQLOLIPQGVTL-----PGPGQLOMQAMPNGVQREFLPLATATAS 2083
Db 1103 -----GETISGCS--KRYVTTPCSTGTCGEYTTA-----ITLVYTTA----- 1138
Qy 2084 TTTTSTTAAGTEGROSKLSIPOMQYHQDILPPAOSSSVGPAKAPQTAQPSARPOQ 2143
Db 1139 -VTTVVTTESSTGTSAGKTTTG--YTKSVPTTYVTTLAPs----- 1178
Qy 2144 TOPSPAPQEVOTQEVOTQTVTSNHPSEKQPHAOSSKROVAAQOSPOSNOCOSVPR 2203
Db 1179 -----APVTPATNAVPTTTTECSAATINAGETTSVCSAKTIVSASAGETTAPSAT-- 1221
Qy 2204 VQSPQTRIRPST--PSQLSPGOOSQOVQT--TSQPIPIQPHSLQIPSGQPOPOV 2258
Db 1232 --TIVYTTAI-PTVITTESSAGTNSAGETTGKTSIPTIYITL-IPGNGAKNNEYV 1287
Qy 2259 QSTQOTLSSGQTLNOVSVSPSPQOLOIQOPQOVIANPOLQOOVYVLSQIOSOVAAQIQ 2318
Db 1288 ATAT-----NPISIKTTS-----QIATTAASASSVAPV 1315
Qy 2319 AAOQGVPOQIKLOLPIQIOQSSAVOHOIONV-VTVQAS 2357
Db 1316 TSPS-----LTGPHLOSAGSAVAITYSVPSISSTYOGAA 1348

RESULT 7
US-09-815-242-12713
: Sequence 12713, Application us/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: CURRENT APPLICATION NUMBER: US/09/815,242
```

```
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12713
: LENGTH: 2344
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 2.8%; Score 401; DB 10; Length 2344;
Best Local Similarity 18.0%; Pred. No. 3,7e-10;
Matches 336; Conservative 271; Mismatches 836; Indels 424; Gaps 57;

Qy 797 EGRKNMDEMMDSDKRCRSPKRIKEPDSKDEKVGSDAKGADQNMDSIKITEK 856
Db 747 EVTRNMSDSVSTSGSTQSOQSVSTKADSOASTSTSGS----- 786
Qy 857 DQVKEHLSDSDPKCKEPEMEYDDDKTESHVNCSSQV-----DVNVSEGFHLRT 910
Db 787 -----IMTSTASTSKSTSVSLSDVSASKSLSTSESNVSSSTSTSLVN----- 831
Qy 911 SYKKTKSSKLDGLIER--RIKQFTLEEKQRLKIKIEGKIGKIGTSTSSKNLSESV 968
Db 832 ---SQSVSSMSGSGVSTSLSDF--ISNSSSTESK--SVSTSTSLRTST 878
Qy 969 IYKAKGCGQSDSMROEPNANNDQPEDLIQCSQSDSVLRMSDPSHTTNKLYPKDRVL 1028
Db 879 LSDSVMSSTGSLSKSLSTSTSDSASTSQSVSDSTSNST-----STSESLSEGST 932
Qy 1029 DVYSIRPETKCPKQNSIENDIEKVDLASRQGEPTKTKGNDFFIDSKLAS----- 1083
Db 933 ESISI-----SNSISNSVASTSKLESQSTISLST-----DSKMSSTESL 975
Qy 1084 ADDIGTLCKNKKPLIGESDITVSSSKALHSVPKSTINDRQATP--LSRAMEGKLG 1141
Db 976 SDSTST-----SDSVSGSLVAGSQSVSTSDSMSTSEMDSSTSGSL 1022
Qy 1142 CD-----SESNSTLENSDPTVSIQDSSEEDMTIVONSNEISIEQFTREQDVE 1188
Db 1023 ASDSKMSVSSMSSTSGSGSTSESLSDSISTSDSDSKSLSTSQSGSTSTSTSSSVR 1082
Qy 1189 VLEPLKCELVGSESTGNCEDRLPYKGTBANGKKPSQOKLLEPRVKNKCSQIKLNTDK 1248
Db 1083 MSE---SQSTSGSMSTSQSSTIS-----TSFSD-----STSDS 1114
Qy 1249 KNNNRESEKKGQRTSFQINGKDNKPKIYKGCLEKIESRVSGVNGVPPVNNIKII 1308
Db 1115 KSASTASSESTSQSVST-----STSGSVSTSLSTSLSNSERTS 1152
Qy 1309 PE-NDIKSLVYKES-----AIRPLNG-----DVIMEDFENRNSSETR-----S 1346
Db 1153 TSMDSSTSLSTSESDSTSDSTSDSTSESLISGSESTISLSESNSTSDSEKSAFAFLS 1212
Qy 1347 HLSSDAE-----GNRDSLETLPTSESDSTQTTTPASCPESNSVQVDEMIET 1399
Db 1213 ESLSESTSESTSESLSTSDSTSLSDSNSRSGSTSL--SNSYSGASISTSTSGAST 1271
Qy 1400 SEVKKVTSPTSEENLSNDFIDENGLPINKENNGE-----SKRKT 1444
```

```
Db 1272 STVK---SESVSTSLSTSTSLSDSTSLSTSLSDSTSGSKNSLSASMSSTSDSISIRKS 1328
Oy 1445 VITEVTMTSTVATBESK---TVIKVEKGDK-----QVVSSTENCAKSTVT- 1487
Db 1329 ESLSASTSLSGSTSESESGSTSSSEKSDSTSMLSMSOSTSGSTSESLSDSTSR 1388
Oy 1488 -----TTTTYTKLSTPSTGGSVDIISVKEQSKYVTTVTDS 1525
Db 1389 LSLASMNOSGVDNSASASASTSTSTSESDOSTSSYSTSOSTSQSESTSTSLSDS 1448
Oy 1526 L-----TTTGGTLVMTVSKFYSTRDYKLMKFSPKTRSGTALPSYRKPFYKSTKS 1580
Db 1449 TSIKSTSGSGSTSTASISGSESESDOSISTSEKSESTSLSDSTSTNSGSA 1508
Oy 1581 IFVLPNDDLKLARKGIREVPYFNNAKPALDIPYPSRPRTFGITWYRLQTVKSLAG 1640
Db 1509 TSTL-----LSNSASASESDSSSTSLSDSTASMQSESDS-----QSTSTSLN 1553
Oy 1641 VSLMLRLMASLRMDMAKVPGGGSTRTESETEITTEILIKRDVGPYGIREFYCI 1700
Db 1554 SOSTSTSMSTIASSEVSESTSESGSTSESTSTSL----- 1595
Oy 1701 KIIPIGIVETKEETPTQPKGLRSALRKRPKPKQGPVILIEVVAEELMEIRA 1760
Db 1596 -----SDSOSTSRSTASGASTSTSTSDSKSTASTSTMTSTLD----- 1637
Oy 1761 FAERYEKEKAQAVEOAKKRLBQOKPTVIATSTTSPSTSTISPAQKVMVAPISGV 1820
Db 1638 -----SQMSLSTSTSTSVSDSTSLSDSVSDSTSDSTSGMSASISLSDST 1688
Oy 1821 TETKM--VITTVGSPATVTPQONKNEHOTFAWVKOGOSNGVVOYQKVLGIPSG 1878
Db 1689 TSTSASEVMSASISDSQSSESVNDESEVSESNSESDSKMSGSTSVSPS--GSLSVST 1746
Oy 1879 TSOQFTSFQPTATYTIPTNSGSGTNSGOVITGPIRPGMYIRPL--OOSTIG 1935
Db 1747 LRKSESVS---ESTISGSGQMSDVSSTSDSSLSYSTQKRSSEVSESDSLSDSK 1803
Oy 1936 KAIIRTPVWVGAPQOVWQIIRQOPVSTAVSAPNTVS-----STPGOKSLTAS 1990
Db 1804 STSTSTSGSLSTSLSGSESVSESSLSDSMSDSTSDSDSLSGSTSLSGSTSL 1863
Oy 1991 IOSSASQPPRPOQGVKLTMAOLITLQGHGNGCL--TVYIGQGGOTTQOLQIPG 2049
Db 1864 SDS-----LSDSKLSGSGQMSGSESTSTSVSDSQSSSTNSGFDMSISA 1909
Oy 2050 LBGPCQQLQAMAPNGVGRFLFTPLATATASTTTTAVSTTAAAGTGROKSLSPQ 2109
Db 1910 SESDSKSTSDSSISGS-----NSTSTSLSTSDMSGVSYSTSTSLSDSISG 1960
Oy 2110 VHODKTLPPAOGSSVGPAPAKAPQTAQPSARPQTPQSPAPQEPVQTPQEVQTT 2169
Db 1961 VSDSS--TSTESLSDMSAQOSTSTASGSLST-----SISLSMASAGLIT 2012
Oy 2170 VSEKAPHTAQSCKPOVAQOSQOSVQOSPVVQSPQTRIRPSTPSQLSPGQOS 2229
Db 2013 VSTSLSTSDSIDSTISIT--SGSQSAVESES-----TSDSTISIDSESL 2066
Oy 2230 TTTSOPPIQPTSLQIPQGOQPOQPOVQOSTQTLSSQTLNQNVSVPSPRQOL 2289
Db 2067 TSTSE-----SLSTNSG---STVSESLSTSGSGST---SVSDSSSTSL 2110
Oy 2290 QPOVIAPVQLOQOVVLQIOQOVAQLOAQOSGVPQOKILOLPIQIOQSAVQTH 2349
Db 2111 GSTVSQDSTSMESMSASISMSGOSISGSTSGSTSTSTSLSGSTHSTSVS-- 2168
Oy 2350 VYTVQAAVQEOQLOR-----VQOLRDQOKKKQOQOIEIKREHTLQASNOSE 2401
Db 2169 ISTNSGMSNSIRHFTSLSTSGMLSLSDNSGMSDSDSVISASBSMSASMSD 2228
Oy 2402 VKK--HNAVIEHLKQKSKMTPAERENQMIYCNVYKILDKIDKEQAKKRKR 2459
Db 2229 TSSSMNSMSKSTSESNLHPSNMSSTSO-----THFTSTSTSEISA----- 2274
```

```
Oy 2460 VEQKRSQONATKLALLFKH-----KEOL--RAELIKRALLDKLOIEVOELK 2510
Db 2275 TNSQSTLSNTSVSS---KHDAEPAQSEERLPDGTGSIQNQNLGLGVMLLVGL 2328
Oy 2511 KIKKEND 2517
Db 2329 KRRKKRD 2335

RESULT 8
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.47JC13
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match 2.7%; Score 391; DB 10; Length 5179;
Best Local Similarity 22.5%; Pred. No. 2.4e-09;
Matches 221; Conservative 91; Mismatches 418; Indels 254; Gaps 33;

Oy 1366 PSTKESDSTQTTTPASCEESNVNOVEDMEIETSEVKKYTSSPITSSEESNL 1425
Db 3460 PTTTPTTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3508
Oy 1426 NGIPLINKENV-----NGESKRKTIVTEVTMTST--VATESKTVIAKEG 1475
Db 3509 --PTPTTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3566
Oy 1476 SSEENAKASTVTTTTVTKLSPTSGSVDIISVKEQSKYVTTVTLSLTTGGT-- 1533
Db 3567 PPTPTGQTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3626
Oy 1534 VTSMTVSKSEYTRDKVLMKFESRPKTRSGTALPSYRKFEVSTKKSIFVLPN 1593
Db 3627 TTTTPT-----PTPTPTGQTPPTPTPTTTT----- 3654
Oy 1594 RKGIREVPYFNNAKPALDIPYPSRPRTFGITWYRLQTVKSLAGVSLML 1653
Db 3655 -----VPTPTPTGCT-----QPTPTPTPT 3673
Oy 1654 WDDMAKVPGGGSTRTER--SETEITTEILIKRDVGPYGIREFYCIKILCIP 1710
Db 3674 ---TTTTVPTPTPTGQTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3720
Oy 1711 TPKEPTPQPKGLRSALP-----KRPEPKQGT-----PVILETVVAEEL 1758
Db 3721 TTTVPTPTPTGQTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3771
Oy 1759 RAFAERVEKEKAQAVEOAKKRLBQOKPTVIATSTTSPRS---STTSTISP 1809
```


Db 1454 KDVAKDELATKANEROKALITQUTADATTEKEQANOCVDAQLTQGNQNIENAQSIDDVNTA 1513
QY 2033 QSGTTGQLOLIPGCVYVLEPBGQOLQAMPNQTVQRFLETPLATATTT----- 2081
Db 1514 KDNAIQALDIPQASTVDKTNARAELLT-----EMOKITTEILINNETTEEKGNDIGPV 1567
QY 2082 -----ASTTTTSTTAAGTEGQROSKLSPOMOVHODKTLPPAOSSSVGPAA 2129
Db 1568 RAAYEBELININATTTGDV--TTAKDPAVOK-----VOOLHNPVKKPRAGKKELOQAA 1620
QY 2130 QPOTAPSAFPOPOPOSPAPPEVOTQPEVOTQTVSSHVSEADPTAHOSSKPOVAAQ 2189
Db 1621 DKKT-QIECTPNASQOEINAKOEVDTELN-QAKTNVDSSTNEVDNAVEKAKAKINA- 1677
QY 2190 SQPQSNVQSGSPVRYVQSPQTRT-----RPSRPSQSLSPQSQSOV-----QTTT 2232
Db 1678 VKTFSEYKKDALAKIDAVYNAKVNEDNSNATSSSEIAEKQKLAELKQTDADONVQATIS 1737
QY 2233 SQPIPIQPHSTL-QIPSQGPQSPQVQSSQTLSQGLNVSQSVSSPSRPOLOIQOPQ 2291
Db 1738 KDLIEVQIHNLDLNDINTTPTGKKSATTDLXAYADQKKNNISADTNATQ----- 1788
QY 2292 QVTAVPOLQOQOVVLSQISQSVVAAQIQAOQSGVPQOIKLOLPQIQOQSSAVQTHQIQNV 2351
Db 1789 -----DEKQQAIRKQVDQNVQTALESINNGV-----DNGDVDAQTQGA 1827
QY 2352 TVQASVQEOLOVQOLRQOQOKKQOQIEIKREHLLQASQSEIIOKVVMHNVH 2411
Db 1828 ALIDAIQVDAV-----KFRANQALIEVKADPTKESIDQSDQLAEKTEALAIKQ 1877
QY 2412 L--KOKKSMTPA-----EREENQMIVCNOVMKYIIDKID-- 2444
Db 1878 IPTQAKGIDATTTAEVAKAKQGLEAFNIDISTEKQAT---EELTALDQIEAGV 1934
QY 2445 -----KEKQA-----AKRKRESEYQKRSKONAT-KLSALLFKKEQLREIL- 2488
Db 1935 NVNADATTEKEAFNALDELISKATEDISDQTTNEMIAIVKNSAL-----EDLKQORIN 1989
QY 2489 ---KKRAL-----LDKDLQIEVOEELKRDJLTKR-----KEKDLMLQAOATV 2527
Db 1990 PEYKKNALAEIREVYVK--QIEIITKNADADASAKETARTDLGRFDFADKLKDTQINAE 2047
QY 2528 AABCPVTVPLPAPAPPSPPPPGVQHTGLSTPLPVASQKRKEEKSSSKSKK 2587
Db 2048 VAEIQONT--IPALAEIVPQNDPDANDFNNGI-----DNNDATANSNAN 2089
QY 2588 KMISTSKETKDT 2601
Db 2090 APPEMTGQPNVSEPT 2103

RESULT 11
US-09-815-242-12967
Sequence 12967, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

QY 1071 GDNFPIIDSKIASADDIGTLICKNKKPLIOEESDTIVSSSKSLAHSSVSPKSTND-RDAMP 1129
Db 664 -TTEKDNGLAVLEODVITTPV---KPOAK-----QDIIQAVTTKKQOIKKSNAS 708
QY 1016 HTTNK-----LYPKDRVLDOVSIRSPETCKPKONSTIENDIEEVSDLASGQPTSKTK 1070
Db 627 QSKKVVQDEYVALVTK-----INDDKNNAIAEINKQTTAAQV----- 663
QY 959 SSKNLSESPV---ITAKKECQSDSRQEDSPANNQDPEDLIQGCSQSSSVLRKSDPS 1015
Db 573 TIINEDANHVKA--NRASQADIDGLV-TKLQALIDNQAAIAELDTKAQEK---VTAQ 626
QY 899 VNAVSEGFHLRTSYKKTKSSKLDGLIERIKOFTLEEKORLEIKLEGGIKIGTSTN 958
Db 514 SDV-VINNAQPEVTLADPESVAVENKMDALQOQVNSQVDNSHYTTASIAEYKKLQQAQ 572
QY 858 QDVKEILSDSDRCKEPEMEVDDDKTES---HVNCQESS-----QVD 898
Db 458 NKVFTTSQGYTTARGHTKKEVLEPDKSLKSYVNVANANIDTPKNIDPNE---KLIFKRA 513
QY 812 KRKCSRSP-----KKIKIEPD-----SEKDEVKGSAAKAGADONEMDISITEKKD 857
Db 400 IGNNGNFGASLKAQDFEYEVTLPOGYTVN--NSLTFTFPNGNEDSTVLKNMVTNVDQNA 457
QY 765 --GYGCHSWISKTHVRFVKLP-GNTNVVYRKSLSEBT-----KNNMDENDESD 811
Db 340 FIDPNTAVADASRTTQKDGKYYSFIDNVGLPSGSHLYKNBDLAPKATNKEYTINTE 399
QY 723 KYKKEKKQEEETMOATVVKY-----TEPVKHQVKKR-----GEERYVT-- 764
Db 664 HRANWIKAVQMGSKPREFALALAILBCAVPVVMLPIWREFLHTRLHR--TSIEREKE 722
QY 291 MQTN-----SKQATNLVIYKD-AKKNTELATVYNAKGTAILFKYPTADRLDLO 339
Db 262 DSTD-----HGNENG-----IKTIVNVKDSBLIINFYT 290
QY 604 EBDHKRRHLAKFCPLPAGEFKNNGSYHGSKVLITSLRLITTOLENNIPSSFLHPNWS 663
Db 544 GKEVLVNSQGEISRLSTKREVIIMKGNINNYFKLQEGKYRVHNOYSTNSFALNRHNR 603
QY 223 GK-IGFNSE-----LRSDMFDKNPNQV-----QAKGVVALGVRVYAN 261
Db 163 PAAPVANAQTOPPNAISNTGEGSINTLTPDDPAISTDENRQDPTVYTDKVGYSLLIN 222
QY 499 PGAGKAGSGSTR-----IITRLRNPDSKLSQKSGO--VAAAHNEAN--KLFE 543
Db 103 NTPAGQGNQADPNNAQAOPGNQATPANQAGQGNNOATFPNNATPANQOPANAPAAQ 162
QY 446 NDSKDAEKNREEDDSLEKSDDKTPDDDPDEGKSEVGD-----FKSEKNGELSES 498
Db 103 NTPAGQGNQADPNNAQAOPGNQATPANQAGQGNNOATFPNNATPANQOPANAPAAQ 162
QY 499 PGAGKAGSGSTR-----IITRLRNPDSKLSQKSGO--VAAAHNEAN--KLFE 543
Db 163 PAAPVANAQTOPPNAISNTGEGSINTLTPDDPAISTDENRQDPTVYTDKVGYSLLIN 222
QY 544 GKEVLVNSQGEISRLSTKREVIIMKGNINNYFKLQEGKYRVHNOYSTNSFALNRHNR 603
Db 223 GK-IGFNSE-----LRSDMFDKNPNQV-----QAKGVVALGVRVYAN 261
QY 604 EBDHKRRHLAKFCPLPAGEFKNNGSYHGSKVLITSLRLITTOLENNIPSSFLHPNWS 663
Db 262 DSTD-----HGNENG-----IKTIVNVKDSBLIINFYT 290
QY 664 HRANWIKAVQMGSKPREFALALAILBCAVPVVMLPIWREFLHTRLHR--TSIEREKE 722
Db 291 MQTN-----SKQATNLVIYKD-AKKNTELATVYNAKGTAILFKYPTADRLDLO 339
QY 723 KYKKEKKQEEETMOATVVKY-----TEPVKHQVKKR-----GEERYVT-- 764
Db 340 FIDPNTAVADASRTTQKDGKYYSFIDNVGLPSGSHLYKNBDLAPKATNKEYTINTE 399
QY 765 --GYGCHSWISKTHVRFVKLP-GNTNVVYRKSLSEBT-----KNNMDENDESD 811
Db 400 IGNNGNFGASLKAQDFEYEVTLPOGYTVN--NSLTFTFPNGNEDSTVLKNMVTNVDQNA 457
QY 812 KRKCSRSP-----KKIKIEPD-----SEKDEVKGSAAKAGADONEMDISITEKKD 857
Db 458 NKVFTTSQGYTTARGHTKKEVLEPDKSLKSYVNVANANIDTPKNIDPNE---KLIFKRA 513
QY 858 QDVKEILSDSDRCKEPEMEVDDDKTES---HVNCQESS-----QVD 898
Db 514 SDV-VINNAQPEVTLADPESVAVENKMDALQOQVNSQVDNSHYTTASIAEYKKLQQAQ 572
QY 899 VNAVSEGFHLRTSYKKTKSSKLDGLIERIKOFTLEEKORLEIKLEGGIKIGTSTN 958
Db 573 TIINEDANHVKA--NRASQADIDGLV-TKLQALIDNQAAIAELDTKAQEK---VTAQ 626
QY 959 SSKNLSESPV---ITAKKECQSDSRQEDSPANNQDPEDLIQGCSQSSSVLRKSDPS 1015
Db 627 QSKKVVQDEYVALVTK-----INDDKNNAIAEINKQTTAAQV----- 663
QY 1016 HTTNK-----LYPKDRVLDOVSIRSPETCKPKONSTIENDIEEVSDLASGQPTSKTK 1070
Db 664 -TTEKDNGLAVLEODVITTPV---KPOAK-----QDIIQAVTTKKQOIKKSNAS 708
QY 1071 GDNFPIIDSKIASADDIGTLICKNKKPLIOEESDTIVSSSKSLAHSSVSPKSTND-RDAMP 1129


```

Db 709 -----LQEDOVANDKIGKIEFKAIKDIDAATTNAQVEAIKT-----KAINDIQOTTP 756
Qy 1130 ISRAMDFESKLCQDSESNSTLNSDPTVYSIODSSEEDMIVQNSNESISQOFPTREODEVY 1189
Db 757 AJTA-----KAAALEEFEEVVOAQ-----IDQAPLNP-----TTNEEVAEA 793
Qy 1190 LEPPLCELVS-----ESTNCEDRLPVKCTE-----ANGKKPSQOKKLEERPYNKCSQJIK 1241
Db 794 IERINAAKVSQKAIETATTTADLERVKNKEISKIENTIDSTQTKDAYNEVQAATARK 853
Qy 1242 LKNTDDKKNNESEKQORTSTFOINKDNKPKIYKCECLKEISESEVAVSGNEPKV 1301
Db 854 AQMAT-VSMATEEVEVAEADAQAQKGLHIOYVKS-----QEVADIK-----SKLIDKI 905
Qy 1302 NNIN---KIIPENDIKSLTVKESAIRPITNGVIMEDFERNSSSETKSHLLSSDAEGNV 1358
Db 906 NAIQOAKVPAAD-----TEVENA-----YNTKQEOIQNSNASTTEEKQAAV 948
Qy 1359 RSLETLPTKRESDSQ-----TTTSPASCPESNS---VNOVEDMEIETSEVKVYSSPIT 1411
Db 949 TE-----LQTRKQOARTNLDANANTNSDVTYAKDNSIAINQVOAATTKKSDARA----- 997
Qy 1412 SEESNLNDPFIENGELPINKNENVNGESKRKTIVITEVTMTSTVATESEKTVIKVEKDK 1471
Db 998 -----EIAOKASERKTAIEAMNDST-----TEEQOAK-DKVDQ 1030
Qy 1472 QTVVSTTE-----NCAKSTVTTTTVTKLSTPSTGSDVLIISVEKQSTVITVTD 1524
Db 1031 AVYTNADIDNNAANDVDNNAKTNEATIAAI-TPDA-----NVKPAKQAIAQVQA 1082
Qy 1525 SLTTGGTILVTSMYVSKESTDQKYLKMFSPKKTSTGAL-PSYKRVYVSTKKSITV 1583
Db 1083 QETAIDG-----NNGSTEEKAAKQOVQTEKTADAIDAATAHTNAEVEAKKA--- 1131
Qy 1584 LPNDLKLARKGIREVYFYNNAKPAIDIMYPSRPFTGITYMYRIQYVKSIAVSL 1643
Db 1132 -----AIAKIEAIOPATTTQDNKEAI----- 1153
Qy 1644 MLRLMASLRWMDMAKVPGGSGSTETSETETITTEIKRRDVPYCIREFYCIKTI 1703
Db 1154 -----ATKANERKTAIAQOTITAEITAANAADVMATQAN 1190
Qy 1704 CFIYVETPKETPTPOKGLRSALPKRPETPKOTGPVITETWAEEL-----ELME 1757
Db 1191 SNEIANSNDVQOAKTGENSEI-----DQVTPVNNKATARKNITAILNKKLOE 1240
Qy 1758 IRAPAEVKEKAQAVEQOAKRLEQKPTVIATSTSTSSSTSIISA-----QXWAV 1812
Db 1241 IQATPDATDEEK-QAADAENANTENGKANOAISATITNAQVDEKANAEEAIAINAVTPKVV 1299
Qy 1813 ABISGSVTTGTMLVLTGVSPATVTFQONKNFHQTFATWVKOGQSN-----SGVQV 1865
Db 1300 KOAAMDEIQOATQTNVINNDQNAATTEKEKAAIQOLATAVDAKNNITAAITDNDNVDA 1359
Qy 1866 QOKVIGIISSTGTSQOTFTSPQPTAVYTIIPNTSGSGTISNSOV--ITGPQIRPKMT 1923
Db 1360 KD-----AGKNSIOSTQPATAVKSNKNDVDAQVTTQMOAIDNTTG----- 1400
Qy 1924 VTRPLQOSTLCKAIIIRPVWVQPGAPQOV-----TOIIRQOPVS--TAVSAPMTV 1973
Db 1401 ---ATTEENNAKDLV---LKAKEKAYODILNAQTTNDVTO--IKDDAVADIGITVADTTI 1453
Qy 1974 SSTPOKSLTSASTSNIOSSASOPPRPOGOVKLTM-AQLTOLNOGHGNGGLVAVIG 2032
Db 1454 KIVANDELATKANEKALIAQATADATTEKEKANOQVQOLQOGNINENASIDVYNA 1513
Qy 2033 OGQTTGQOLIPQGVTVLPQGOQLMAQAMPNGVQORLETPPLATTAT----- 2081
Db 1514 KONAIOAIDPIQASTDVKTNARAEILT-----EMQNKITEIILNNETNEEKNDIGV 1567
Qy 2082 -----ASTTTTIVSTTAAGTGEOQSKLSPOQOVQODKTLPPAOSSVCPAPA 2129
Db 1568 RAYEBGLNINNAATTGTV--TTAKDTAVQ-----VOOLHANPVKKPAGKRKELDQAA 1620

```

```

Qy 2130 QOQTAQPSARPOQOTQPSAPQEVQTOPEVQOTQTTVSSHVSEADQPTHAQSSKPPVAAQ 2189
Db 1621 DKKT-QIEQTPNASSQOEDINDAKQOEDVTELN-QAKTVVDOSSTNEVDNAVKEGAKINA- 1677
Qy 2190 SQPSNVQSGPVRYQSPQTRI-----RPSPSQSLSPGQOQV-----QTTT 2232
Db 1678 VKFSEYKKDALAKIEDAVNAKVENADSNASTSSEIARQKLAELKOTADONVQATS 1737
Qy 2233 SQPIPIQPTSL-QIPSGQPSQPSQPOVQOSTQTLSSGOTLNOVYSSPSRPOLOIQOPQ 2291
Db 1738 KQDIEVQIHNDLDNINDYITIPGKKESATTDLYAVADQKKNISADTNMTQ----- 1788
Qy 2292 QYIAPVQLOQOVVLSQIOSQVVAQIOAQSGVPQOIKLQLPQIQOSSAVQTHQIQNV 2351
Db 1789 -----DEKQOAIKQOVQNNQTALESINGV-----DNGDVVDALTOGRA 1827
Qy 2352 TYQASVQSOLOQVQOLRQOQKQKQOQIEIKREHTLOASNOSEIIQOVYKNNAVIHH 2411
Db 1828 AIDAIOVDATV-----KPMQALEVKAEDTKESIDQSDQLTAEKTEALMIXQ- 1877
Qy 2412 L--KOKSMTPA-----EREENQRMIVCQVAKYILDKID----- 2444
Db 1878 IYDQAKQGITDATTAEVKAQAQGLEAFDNIQIDSTERQKAT---EELETALDQIEAGV 1934
Qy 2445 -----KEKQA-----AKKRRESEVEQKRSKQNAF-KISALLFHKQELRAEIL- 2488
Db 1935 NVNADATTEKEKAFNALDELISKATEDISDQTNIAITVKNAL-----EQLKQORIN 1989
Qy 2489 -----KRAL-----LDKQDQIEVQELKRDKIK-----KENDMQLOAQATV 2527
Db 1990 PEYKKNALAIAREVVK--QIEIINKNADADASAKELIARDLGRFYRFDKIDKTQTNAE 2047
Qy 2528 AAPCPPTVLPAPAPPPSPPPGVQHTGLSTPLTLPVASQKRRKEEKSSSKKK 2587
Db 2048 VAELOWVT--IRAIEIYQNDPDANDTNNGI-----DNDNATANSNAN 2089
Qy 2588 KMISTSKETKKDT 2601
Db 2090 APWENTGCPNVSET 2103

RESULT 12
US-08-681-219-32
; Sequence 32, Application US/08681219
; Patent No. US20020058607A1
; GENERAL INFORMATION:
; APPLICANT: Takaaki Sato and Junn Yanagisawa
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,219
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JKM
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-681-219-32

Query Match 2.3%; Score 336; DB 8; Length 2843;
 Best Local Similarity 17.1%; Pred. No. 3e-07;

Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;

QY 387 EIRHMDITFDLTKKAGSKSFLA-ANELESTRAKGDDIDNYSPEPEKDKN---442
 DB 802 DTRHRDNRSDNNT---GMATVLSPLNTTVLPSSSSSGSLDSRS---EKDNLER 854
 QY 443 -----ETENDSKDAEKNREFEEDS-----LEKSDDKTPDDPEQK-----480
 DB 855 ERGIGLGNYPATENPGTSKRGLQISTTAQIAKVMEVSAIHTSQEDRSSTTELHC 914
 QY 481 -----SEVGEF-KSEKSN-----GELSESPGAG 502
 DB 915 VTDERNALRRSSAAHTHSNTNTNTEKSENSRTCSMPYAKLEYKRSSNDLSNVSSSDGYG 974
 QY 503 KGASGSTRITRLRNPDSKLSQLSK---QQVAAAHAEANKL-FKEGKEVLVN-----551
 DB 975 KRQGMPSIESYSEDESKCSGYCPADLAHSHANHHDDGELDPIINSLKYSDE 1034
 QY 552 --SOGELSRSTKKEVIMKGINNVEKLGQEGKYRVYHNOYSTNSPALNKHQREDHK- 608
 DB 1035 QLSNGQSPQSONERMARPKHIEDETIKOSEQRQR---NOSTT--YPVYESTDDKHLKF 1089
 QY 609 RRLIAHKFCLTPAGEFKKNSVHSGKVLITSLRLITTOLENNIPSEFLHPNASHRAM 668
 DB 1090 QPHFGQCECVSPYRSGANS-----ETNRVGS-----NKGINO 1123
 QY 669 IKAVMQSKPREFALALALIECAVKPVMLPIWREFLGHTRLHMTSIEEEKKVKKE 728
 DB 1124 NVQSILCOED-----DYEDDKPTNINSE 1145
 QY 729 KQOEEETMOQATWVKYTPPVKHQVKQGEYRVTYGQSWISKTHYRVPKPLRGNT 788
 DB 1146 RXSEEE-----QHEEERPTNY 1162
 QY 789 NVNYRKSLEGTKNMNDENDES-----DKRCSRPKIKITIEPDEKD 831
 DB 1163 STATNEE---KRHDOPIDYSLIKATDIPSSOKQSPFSKSSSGSKTEHMSSESENT 1218
 QY 832 EYKSDAAGADQNDMDISKTEKKDQDYVELDSDS-----DKP-CKEERPA 877
 DB 1219 SPFSSNAKR---QNOIHPSSAQSRSGQPOKAATCKVYSINQETIOTYCVADPIICFSRCS 1275
 QY 878 EYVDDMKTSHVNCQSSQ-----VDVNVNSEGFHLNT-----SYKKTKSS 919
 DB 1276 STLSLSSAEDELICQNOTTEADSANLQIAIEIKETIGTSAEDPVEVPAVSQHPRTKSS 1335
 QY 920 KLDGLERRIKQFTLEKORLEKIKLEGGIKGIGCTJSTSSKN-----LSESVI-TRAK 973
 DB 1336 RLOG-----SSLSESAHKAVERFSSGAKSPKSGAQTPKSPHYVOEPLMFSRGT 1388
 QY 974 EGCGSDMKROEOPNANNQDPEDLIOGCSQSDSVLRMSDPSTHTTKLYPKDVLVDVSI 1033
 DB 1389 SVSLSDFSRSTIASSVOGEF-----CSGMSGIIISPDLDPSPQOTMPPSR- 1435
 QY 1034 RSEPTKCPKONSTENDIEEKVSDLASRGQPTKSKTGKGNDFITDSKLASAD----- 1085
 DB 1436 --SKTTPPPPOATQTKRE-----VPKNKAPTAKEKRESG---PKQAAVNAAYORVOY 1481
 QY 1086 --DIGFLI-----CKNK-----KPLIOEESDTIV-----SSKSALHSSV 1118

DB 1482 LPDADTLHPATESTPDGFCSSSSLALSLEDEPFIOKDELRIIMPVQENDNGNETSEIQ 1541
 QY 1119 PKSTNDRADATPLBRADPEKLCQDSESNSTLNSSQTVYSIODSPEDMIVQNSNISIE 1178
 DB 1542 PKSENNQ-----EKEAETIDSEKD--LLDSDDD----- 1570
 QY 1179 QFRTRODVEVLEPLKCELVGSGSTGNCEDRLPVKGTTEANGKAKPSOOKLEERPVKCSQ 1238
 DB 1571 -----DIEILE-EC-IISA-----MPTKSSR-KAKKPAQDTASKLPPVARKPS 1610
 QY 1239 QT---KLKNTDKKNNEENRESEKKQGTSTFQINGDKNKIKYLKCECLKIEISESVYSG 1295
 DB 1611 QLEPVYKLLPSONRLOQKHVSFTPG-----DdMPrvY-----CV-----EGPTIN 1651
 QY 1296 NVBPKVNNIKIIPENDIKSLVAKESAIRPFINGDIYMEPFNRNSSEFKSHLLSSDAE 1355
 DB 1652 STATSLSDLTIESPPNL-----AAGGVAGGAQSGEFERKDTIPIEGR--STDEAQ 1701
 QY 1356 GNYRDSLETLPSYTES-----DSTQTTTPSASCP 1384
 DB 1702 GGRKTSV-TIPELDDNKAEBGDIACINSAMPKSGHKPFVKKIMDQVOQASASSAP 1760
 QY 1385 ESNVAVQVEDMELETSEVKKVT-----SPTSEBSNLN 1420
 DB 1761 NKNQLOGKK--KKPTSPVKPIQNTERYTRVRKNADSKNNJNAERVSDNKSQKQNLKN 1818
 QY 1421 DFTDENGCLPIKNMENYNGE-----SKRR 1443
 DB 1819 NSKDFNDKLPNNDRRGSAFDSPPHHYTPLEGTPICFNSDLSLDDPDDVDLSREK 1878
 QY 1444 TVITEVYTMSTVATKESKYIKVEKDKQTVASSTENCAKSTVY-----PTTTTV 1493
 DB 1879 ---AELRAKENESBAKVTSHTEITSNOOSANKTQAIKQPINRQKPILOKQSTFPQ 1935
 QY 1494 TKLSTPSTGSDVIDIYVKQDSKTVVTTVYDLSITTTGGTLYVMTYSKESTDKYKLMK 1553
 DB 1936 SSKDIPDRGAID--EKLOFAIENTPVCF--HNSLSLSLSDIDQENNEKEPIKE 1989
 QY 1554 FSRP-----KKTSGTALPSYRK-----EYTKSTKSIYFLPNDDL----- 1589
 DB 1990 TEPPDSQGEPSKPOAGYAPAKSFHVEDTPYCFERNSSLSLSDSDLLQECISSAMPK 2049
 QY 1590 -KILAKKGIREVPYFNNAKP-----ALDIWPPSPRPFGIT-----WRY 1630
 DB 2050 KKKPSRLKGDNE-----KHSPPRMGIIIGEDLTLLDKIQRPDSHGSLSPSENDMKA 2103
 QY 1631 RIQTVASLAGVSLMLRLMASLFWMDMAKVPPGGSTPTETSEETITTEILKRDVCP 1690
 DB 2104 IOEGANSIV-----SSLHQAAAACL-----SKOASDSD--STLSKS--- 2140
 QY 1691 YGIRFEYCIRKILICPIGVETPEKETPTPQKGLRSALPKRREPRTKQTPVITETVAE 1750
 DB 2141 -GI-----SLGSPHLHPIDQEKFTSNKGR--ILNPGKST-----LETKKIE 2182
 QY 1751 EF-----LELWIRAFARVEKEKAQAVEQAKKRLQOKP-----TVIAT 1791
 DB 2183 SESKGIGKQKVKYKSLITGKVRNSSE-----ISGOMKOPLOANMPSISGRRTMIHI 2233
 QY 1792 STTSPTSTSTSPAKQWVAPISGSVYTGTKMVLITTVGSPATVTPQNNKNHQTFT 1851
 DB 2234 PGVRNSSSTSPVSKGPPKTPASKSPSEGGQATTSPGAKPSV----- 2278
 QY 1852 WVKOGSNSGVVOQCVLGIIIPSSYGTGQQTFTSFQPTAVVTLRPNTSGSGGTTSSNQ 1911
 DB 2279 -----KSELSPARQTSQIG-----GSSK-----APSRSGSRDTPS- 2310
 QY 1912 VITGPIRGMVYIRPPLQOSTLGKAIIRT PVNVOGP-----APQVMTQIIRGQPVSTAV 1967
 DB 2311 -----RPAQOPLSPRIOSPGRNS-----ISPGRNGISPPNKISOLDPTSSPSTA- 2354
 QY 1968 SAPNTVSTPGQSKLTS-----ATSTSNIOSSASQPPRQOQVLTMAQLTOLQ 2018
 DB 2355 ---STKSSGSGKMSYTSPPGRMSQONLTKOTGLSKMASSIPSEAS---KGLNOMNN 2406


```

Db 1037 LMOODPKSVRL-----PVSQ 1052
QY 2652 STEDAMTVLTPTLTK---DYEGIKRVLBSLOAHKMAWFELEPVDPNDAPDYGYIKKEMD 2708
Db 1053 NVHPPRGFLNPDSDQMPMQSGSVPMVSLQG-----PASVPPSPDKO---RMPMP 1100
QY 2709 LAT-MEERVOQRYYEKLTEFVADMTKIFDNCRYYNPSDSPFYOCACVLES 2757
Db 1101 VNTPLGNSNRKMYQESPO-----NPSSSPLAEMASLPEA 1135

```

Search completed: November 20, 2002, 16:45:59
 Job time : 88.712 secs

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 3674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.0%; Score 432.5; DB 4; Length 2137;
Best Local Similarity 17.2%; Pred. No. 3e-17;
Matches 395; Conservative 355; Mismatches 999; Indels 547; Gaps 77;

QY 394 ITEDLTNKAQNSKFLAANEI-----LESIRAKGIDIDNYSKPEETEKD- 440
DB 149 MSSSVTNTDSSERKAGISQSSSETSNOSSELTNYASTDHESTTNDNTAQQQNKSSN 208
QY 441 --KNETENDSKAKNEEFEDQSLKSDDKTPDDPEQKSEVDFKSEKSNGLSES 498
DB 209 VTSKSTQSNSSSEKSNISNLQSIETKATDSLATSEARTSTNOISMLTSTSNQSSPT 268
QY 499 PGAKGKSGSRITRLRN-----DSKLSQKSOVAAAAHEAKLF-----KEG 544
DB 269 SFAMLRFSRPTVNTMAAPTSTTTTSLTNSVYVKNDFNEHMLSGSATVDPKGTG 328
QY 545 KEV-----VNVNQGELS---RLSTKEVIYMKGNIN---NYF----- 575
DB 329 IATLTTPAISKGALISLNRDLDSNRSFRFTGKVALGNRYBESPDGYTGCGIGTAFASPG 388
QY 576 --KLQGEG-----KYRVYHNOYSTNSFALNKHQRE-----D 605
DB 389 PLQIGKREGAVGIGINNAPFKLDYHNTSTPKSDAKADPRNVCAGGAFVSTG 448
QY 606 HD-----KRLHAKFCLTPAGE-----FKNGSVHSGKVLITLTLITQLENNIPSS 655
DB 449 RGNMATEASSAARLNVQPTDNSFQDFVIDYNGD---TKVMTVYAGQTFTR----- 497
QY 656 FLHFNMAASHRANWTKAVOMCSKPREFALALALECAVPVYMLPIWREFLGHTRLHMTS 715
DB 498 -----NLTDWIKN-----SGTTFISLMTASTGCAKNLQOVQCFTEYTESAVAKRY 545
QY 716 IEREE-KEKYKKKEKKOEETMOATVWVYKTFPVKHQV--WKQKGEERYT-----GY 766
DB 546 VDAMTNGDIIIPKTIACEVDATV-----IDKOLNMLKNSGYSYVTDALQNSNY 595
QY 767 GGMWMI-----SKTHVYRFVPKLPQNTNVNRYKSLGKKNMDEMDSMRKCSRS 818
DB 596 SETSGTPTLKLJNSSQVTVYKF-----KDVQGPQISVD-----SQTREVGKT 637
QY 819 PKRKLI-EPDSEKD-----EYKSDAANKAQD-----NEMDISKITEKKDDVVEL 864
DB 638 INPTITTTTDSKDLVLTVTYGLPSGLSFDQTTMTITGTSPEVGTITYT-----VNT 690
QY 865 DSDSDCKPEKPEMVDMDKTESHVNCQESSOV-----DVNVWS 903
DB 691 DATGNVTSKQFTITQDIPSVNVTPSQASEVFTPIPIITATDNGSKVVTHTVGLP 750
QY 904 EGFHLRTSYKKKTKSSKLIDGLERRIKQFTLEKQORLEKTLLEGICIGCTSTNSNNTL 963
DB 751 QGLKFDASTNSIVGTPQIGTNTTITESTDASGNKTYTKIYV-----VTRNSASDS 802

QY 964 SESPVITKAKEGCO-----SDSMROBOSPANNDOPEDILQGCQSQSDSVLRMSDPSHTT 1018
DB 803 TSTISVNSVSTISNSISLSDSVKASQSLSTKSNLSLSTL-----SASTSNSTIQASEASTS 860
QY 1019 NKLYPKRDVLDVYSIRSPETKCPKQNSIENDIEKYSDLARKQEPYKSKTKGNDFFIDD 1078
DB 861 KQL-----SFSASTSTSDASASARKSESTSKSTS-----LSE 893
QY 1079 SKLASADDITLLCKKPKPLQIESDPTIVSSKSAHLSVPKSTNDPDAPIPLSRAMDFEG 1138
DB 894 STSTVSASVSTSES-----ASTSTVSASTSTISDSTSTSDASATKASEAS-TS 948
QY 1139 KLGDSESNSTLENSDPTVSIDOS--SEEDIMVQNSNESIEOPRTREQDVEVLEPLKCEL 1197
DB 949 KLSESVSTSTSDASASTSVSDNSKASTSLKSTSVSDSTSTSD----- 997
QY 1198 VSGESTNCEDRLPVKGTANGKKRPSQKKLEBRPVKCSQDQIKLNTTKNNENRESE 1257
DB 998 --SASTSTSE-----SESDSASTSLSESTSTSVSDSTS--TSTSDASASASESE 1043
QY 1258 KKGORTSTFQINGKDNKPKIYLKGECEKTESERVSGNVEPKVNNINKIIPENDIKSLT 1317
DB 1044 SNKSTST-----LSE--STSLSGSTSTAST-----SDGASTS 1074
QY 1318 VKESAIRPPIINGDIVIMEDFERNSESTKSHLSSDAEGYRDLSTLPTSKESDSTQTT 1377
DB 1075 TSESE-----SDSTSTSLSESTSTSLSGS--TSASTSDAST--STSESDSTSES 1120
QY 1378 TPSASCPESNVNOVEDMEITSEVKRYVTSAPLTISEESN-----LSNDFIDENGL 1428
DB 1121 T--SLSESLSTSVSDSTASTSE--SASTSTSEESNSASTSLSGSLSTISDSTST 1173
QY 1429 PINKENYV--GESKRAKVIPEVTMTSTVATESKTVLKEVGDKQVWVSTENCASKTVT 1487
DB 1174 STSDSASTSTSESDSTSTSLSTSTSLSDSTSTSEASTSTSESDSTSTSLSE 1233
QY 1488 TTTTTVTKLSTPSTGSGVDI--ISVKEQSKTVV-----TTVYDLSLTGGLTVMTWS 1540
DB 1234 STSTSVSDSTSTSDASTSDASTSVSDSEASTSISESLSTSVSDS--TSTSDASTSASTS 1292
QY 1341 KEYSTRDKVKLMKSRKRTKPSGALPSYRKFFVYKSKTKSTFVLPNDLKKLAKGSGIRE 1600
DB 1293 ESDSTSESTSLSEISSTSVSDSTSA--STSDASTSTSESE-----ESDASTSLSG---- 1341
QY 1601 VPYFNVAKPALDIMPYSPRPFGITWRYRLQYVKSSTLAGVSLMLRLMLASLRFMDMAAK 1660
DB 1342 -----STSTSLST----- 1348
QY 1661 VPPGGSTRRETSETEITTEILIKRDVGPYGIREFYCIRKTIICPIGVPTPKETPTPQR 1720
DB 1349 -----DSTSTSDASTSASTS-----ESDSERASTSL 1375
QY 1721 KGLASSALRPKRPETPTQGTGV-----IIFTWAAEELMELMELRAFAER 1764
DB 1376 SGTSTSTSLSDSTSTSDASTSVSDSNASTSLSGSLSTSVSDSTSTSDASTSAST 1435
QY 1765 VEKKAQAVEQOAKRLEQOKPVIATSTSTPTSTSTISPAQWVWAPISGSVTTGTK 1824
DB 1436 SESD-----SRRASTSLSGSTSTISDSTSTSDASTSTSVSE-----SNSTSTSIS 1484
QY 1825 MVLTTKVGSPATVTFQONKNFHQFPAFWKOGOSNGVOYQAVLCIIPBSGTSGOQTF 1884
DB 1485 ESLSTSV--SDSTST-----STSDASTSTSVSDASTSSSESV-----STSDSTSES 1532
QY 1885 TSFQPRATVIT-----RPNISGSGGTTNSQVITGQIRGGMVITPPLQOSTLGKAIL 1939
DB 1533 TSTSDASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSDASTSASTSESDASTSAST 1589
QY 1940 RTPVWQPGAPQOQVMTQILIRQGPVTSVAPNTVSPGQK-----SLTSASTSNISQASAS 1996
DB 1590 SSSSESVSTSVSDSTISASTSEASTSTSVSDSNASTSLSESTSTSLSDSTSMSTSDAST 1649
QY 1997 QPPRPQOGQVKLTMAOLTQLTQGHGNGGLTVVIOGGQGTGQLQLIPQGVTVLPGFGQO 2056

Db 1650 STSESDSDASTSLSDSTSTSTSESTSTSTSVASNSTSL--DSKSTSLSDSTST 1707
QY 2057 LMOAAMPNGTVOFELTPTLATTTATTTTAAAGGEOGROSKLSPQMVOHDKTL 2116
Db 1708 STSESGSTSTSE--DSDASTSLSDSTSTSDSTSTSDASTSMVSDSNRASTSL 1765
QY 2117 PPAOSSV--GPAKAPQTAOPSPARPQTOPSPAPQEPVOTQPEVOTQTVSSHVPSA 2174
Db 1766 SDSTSTSVSDSTSTSTSEASTSTRESESTSTSLSESTSTSVSDSTSTSD--SAS 1822
QY 2175 QPTHAOSKPYQAASQPOSNVWGSPVYQSPOTRIRPSTPSQSLSPGOOSVOTTSQ 2234
Db 1823 TSTSESDSSESTSLSEST--TSVSDSTSTSTSTSTSTSTSDSASTSLSGST 1877
QY 2235 PIPIDHTLQIPSGQPOQPOVQVSSSTOTLSSGOTLNVSVSPARPQIQPOQPOVI 2294
Db 1878 STSVSDSTSTSTSTSTSTSESDSDASTSLSGSTSTSDSTSTSDASTSTSE-- 1935
QY 2295 AVPOLQOQVVLSQLSQVYVAQIQAOQSGVPPQIKLPLQIQOASSAVOTHQIQNVTVQ 2354
Db 1936 -----SASTSTSVSESDSESTSVSESS--TSVSDSTSTSTSEASTSTSE 1979
QY 2355 AASVQOLQVQOLRPOQKQKQOQIEIKREHQLASNOSETIQOYVKNHNVIEHLKQ 2414
Db 1980 SESTSESS-----TSVSESSSTSTSDSSTSTSTSTSTSTSTSTSTSTSTSTSE 2032
QY 2415 KKSMTPEERENORMIVCNOVMYILDKIDKEKOAKKRRKESEVOKKRNATRLSA 2474
Db 2033 DRIYV--QSKNTNML--NKTGK---DYDLQQRGTSEQJNETQSNQADNHSNNL 2083
QY 2475 LFKHKEQLRAELTKR 2490
Db 2084 L--HONRLQDKVYKQ 2096

RESULT 3
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.9%; Score 410; DB 4; Length 3696;
Best Local Similarity 17.7%; Pred. No. 1.5e-15;
Matches 464; Conservative 427; Mismatches 1027; Indels 704; Gaps 108;

QY 311 OKNKPYIRHEPIGYD-RSRKRYWFLNRR--LLEEDTENENKKIYVSTKYVLAELIDC 367
Db 569 QVNKDIT--PSNYTLASVYNNKYLKERAGTVLDEETNMP----- 606
QY 368 LKQDVEALCKILEMREIHRHMDITDLDNKKARG-----SNKSLAANEELIEST 421
Db 607 FQORSQOTQIDDLHELQTLTLNRVSASREINDKAQEMTDVAVDSTELTTEEDTLVDO 666
QY 422 RAKKGQIDVKSPE-----ETEKDNETENDSKAEKKREEFEQSL 464
Db 667 ENHKNEISNNIDELTDQOVERYKAGLHTLLESDTPHPVTKPNAROVVNNRAD-QOKTL 725

QY 465 KDSDDKTPDDPE-----QKSEVDFKSEKNGELSPGAGKAGSSTRITR 514
Db 726 RNNHEATTEQNALINQVEAHSSDALIAKIGAEFTDVTYNARD-----NCKILATD 777
QY 515 LRNPDSKLSQKSOQVAAAHAANKLEKEGEVLVNSQGEISRLS-----TKKEVI 566
Db 778 VNPPTK-----KAEARAAYVNSANSKIKD-----INNTOATLDERNDAIALVNSKPEAI 838
QY 567 MKNINNVFKLQGEKRYVNHQSTNSF-----ALNKHQHRDHDKRLRLAH 614
Db 829 Q--NINT--AOGNDVTEAONN--GNTTIOVPLTPYKRONATATTINAKADEQRLIQAN 882
QY 615 KFLTPA--GEFKWGSV-HGSKVLITSTRLTITOLENNIPSSFLHPMASHRAMWK 670
Db 883 NNATTEKADAEKRVNAVATTANQNTINATINDVDAQTT-----GSGEII 929
QY 671 AVOMCSKREFALALALECAVPRVYMLPIWREFLHTRLHRTSIEREKEVKKKEK 730
Db 930 AISPATKIKEDARA-AVEAKAIAQNOQI-----NSNNMATTEKEKEDALNOVEAHK 978
QY 731 QEEETMQCATVWKYTFPRVAKHOVWKQGEERYVTGCGSMWIKTHYRVVPLPGNTV 790
Db 979 QAAIATITNOA-----QSTQVSE-----AKNNGINTI 1005
QY 791 NYRKSLEGTNNNDENNDESDKRCRSRPPKIRIEPSEKDEYKSDAAKADQONENDIS 850
Db 1006 NQOPNAPVKNKNTKTILEQGNKKS--AIAGTPATTEE-----KQEAVS 1049
QY 851 KITEKKDQVKEILSDSDPKCEEPMEVDDMKTESHVACQSSQVQVYVNSEGFLRT 910
Db 1050 AVSOAVTNGITIHINOANSNDVDQE--LSNAEQIITHTNVVQKRPQ-----ARQALYAKT 1103
QY 911 SYKKTKSSKLDGLLERRIKOF--TLEEKORL--EKI-----KLEGGIKGIGTS 956
Db 1104 NEKOSAINSDNEGTIEKKOAIQSLNDAKYLABQITQAASNONVDNALIGTISNISKIQ 1163
QY 957 TNSKLSSESPVITTKAKEGQSDSMROEOPNANNQPEDLIQCSQSDSVLRMSDP 1016
Db 1164 TNETK--KQARDQVQKQKEABELNSPHTQDEKQALRLQAKFTALNDINOAQ 1220
QY 1017 TTKKLPKPRYL-----DDVSTRPPTKCPKONSINDIEEKYSLASQGEPTKS 1067
Db 1221 TNOV--DTALSGIONIONTOVNNRK--KQEAKTTINDIYQHQKOSIONNDATTEE 1274
QY 1068 KTGNDPEFIDSKLASADDIGTLICKKKPLIOESPTIYSSKSALHSSVPKSTNDROA 1127
Db 1275 KEVANN-----LVNASQ--QNVISKIDNATTNNQIDGIVSDGROSINATITPOTSIRNA 1326
QY 1128 TPLSRAMDFEGLGCDSESNSTLENSSDTVSIO--DSSEEDMIYVONSNEISGEPTRE 1184
Db 1327 -----KNIDIKAKDKIKIQRINDATDEE--IQEANKIEE----- 1361
QY 1185 QDVEVLEPLKCELVSGESTGNCEDRLPYKGTENGK--KPSQOKKLEPRV--NKCSD 1238
Db 1362 -----AKIEAKDNIOBSTRDQVNEAKTNGINKIENTIPATVYKSEARQAVONKANE 1413
QY 1239 QIK-LKNTTDKNNNENRESEKKGORTST-----FOINGKDNPKI-YLKGCELEKTS-- 1288
Db 1414 QIMHIQITPATNEKEQAL--NRVSAELARYQAQOINAEHTTQGVKTTIDDDITISIRI 1470
QY 1289 -----ESRVVSGNVEPRVNNINKIIPENDIKSLTVKESAIRPTNGVIMEDNERNSS 1343
Db 1471 NAQVKEKESARNAIEQATQOTQFINNNDNATDEKEVA-----NNLVIA--TKQSLD 1522
QY 1344 TKSHLSSDAE-----GNRDSLETLPSTKESDSTOTTTPSASCPESNSVQVDEMIET 1399
Db 1523 NINSLSSNDVENAKVAGIENIANVLPATVYKSKAKKDIDOKL--AQOINQIOTHOTAT 1579
QY 1400 SEVKVYTSPTSEESNSLNDFTDNGPLINKNENVNGESKRYVTEVTTMTSYATE 1459
Db 1580 TEER-----EAAIQLANOKSNEARTAIQNEHSHNGVAQAK----- 1614

```

QY 1460 SKTYIKVE-----KGDKQVYSSPENCASGVTTT-----TTVTKLSTPSTG 1503
Db 1615 SNGHEHELYMPDAHKSDKOSIDONKYNSITNTTTPATDEEKOKALDKIAKDAG 1674
QY 1504 --SVDIISVEQ--SKTVVTTTVDLSLTGTLVTSMTVSKREYSTRDKYKL-MKFSRP 1557
Db 1675 YNKVDAQOTNOQVSDAKTEADITLTN-----IQANVAKKPSAR--VELDSKPEDL 1722
QY 1558 KTRRSGLALSYRKFFVTKSKKSLFVLPN---DDKLKLAKG-----GIREVP 1602
Db 1723 K--ROIATN---ATEEKKODAIORLNGKREVKMLINODRRDNEVECHKNGLOELE 1776
QY 1603 YFNNAKPAIDIMPYPSRPFGITWRYRLQTVKSLAGVSLMLRLMASLRMDMAKVP 1662
Db 1777 TIHAN-----PFRKSALOELQI--KFISQTELI-----1803
QY 1663 PGGSSTPSTSEITTEITTEIIRKRDVPGYIRREYCIKRIICPIGVEPTEKPTPQK 1722
Db 1804 ---NNKDATNEEKDEAKRL-----EISKKTITINIQATNNQVDMKNDG 1848
QY 1723 LRSSALPKRPETPKQGPVLIETWAEELMEIRAFERVEKEKAQAVEQAKRLE 1782
Db 1849 MNEIA--TIPATYIKTDATDAIDKKEQOVTI--INGNDATDEEKAERKLVEKAKIE 1904
QY 1783 Q-----OKPVIATSTSPSTSTSTISPAQVAV---APISG 1817
Db 1905 AKSNITSDPEREVNKAKTNEKINNIQSTQTKTNAKEINDKAEOLQIINTP---1961
QY 1818 SVTTGKMLVLTQVSPATVTFQONKNPHOTFATWKGOGNSGVVOQKVLGIPSSST 1877
Db 1962 DATEEKOEAIRNAGIAQAIQINNAHST--OEVNESKINS--IATIKSVOPVNIKPT 2018
QY 1878 GTSQOTTSFQPT-----AT-----VT-----IRPNTSGSGTTSNSQ 1911
Db 2019 AINSLETOBANOKTLIGNDNATDEKEAKQVTOKLINEQIOKIHSTODNOVDNKAQ 2078
QY 1912 VITGPOI-----RPGMTVIRTPLOOSTLGAIIIRTPVWVQCAPQOVMTOIIRGQPV- 1963
Db 2079 AITAIKILINAMAHKRQDAINILTMIAESK--KSDIRAN--QDATTEKKTAL--QSID 2130
QY 1964 STAVSAPVTYSTPGOKSLTSATSTSNQSSASOPPR-----POGGVKLTMAOLTOLO 2018
Db 2131 DTLAQRNNING-----ANTMALVDENLEDKOKLORIIVLSTOTKQAKADIAQAI--- 2181
QY 2019 GHGNGOGLTVIIOGOGTGOLOLIPQGVTLRPGCOOLMAAPNGVQFRLTPIATP 2078
Db 2182 ----GQQRSTIDONAMATTEKQALELNETNGVNDRIQALALANQVDERKNILLETI 2237
QY 2079 ATT-----ASTTTTVSTTAAGTGEOROSKLSPOQVHODKTLPPAOS 2121
Db 2238 RNVEPIYIVKPRANEIIRKKAEOQTTLINQODATLEEKQIALGLKEVKNEALNOVSOA 2297
QY 2122 SSVGRPAK--AOPOTAPQSPARPOPOPOSPAPEVO--TQPEVQI-----2162
Db 2298 HSNNDVKIAENNGIAKISEVHEPITIKRNAKOEIEDOASQSIDITINANKSTNEKSAI 2357
QY 2163 -----QTVSSHVSEAPOTHAOS-----2182
Db 2358 DRVNAVAKIDAINNTNATTLOLVMDAKNSGNTSISQILPSTAVYTNALALASEAKKNNA 2417
QY 2183 ---KPOVAASOPQSN-----VQSGPYRVOSPS 2208
Db 2418 IIDQPNATAEKEERANKYDRLQEADANILKAHTTEDEVANNIKQAQONINAVQEVIAK 2477
QY 2209 QTRIRPSTPQSLSPGQSOVOTTTISQPIPIQPHISLOIPSGQOQSOQVQOSSQOTLSSG 2268
Db 2478 KONVK--NOLNPFIDOKKILINT-----PDATLEKKAANRLLOLVNLSTDELIANV 2528
QY 2269 QTLNOVSVS--SPSRPOLQIQOPQOVIAVPOLQOQVYLSIQSQSVAVO--IOAQOSGV 2324
Db 2529 DHNNEVDALDKARKIE-----ELVPOYSKKRDLVNLMIQEFNFNQTOEIQENDQAT 2580
QY 2325 PQOIKLQPLQIQO--SSAVQHOIQNVVTVQAAVQEOQLQRVQOQLRDOQKKKKQOQIE 2381

```

```

Db 2581 NEE--KTEALINKINQOLLNOQAKVINIDQASNKDVSAKTR--SIDIEQIQHPQTK-----A 2633
QY 2382 IKREHILQASNOSEITQOKOVAMKH--NAVTEHLKOKSMTPAREEENQRIQVQVKYTL 2440
Db 2634 TGHHRILNEKNO--QOASTIAHPNSTIE-----EKQESAKL--OEVLKKA 2676
QY 2441 DKIDKEEQAKRKREESVEQKRSKONATKLSALLFKHKEQULRAELKRALLDQOI 2500
Db 2677 AKIDKQOTNDVAKETVYVNGIAELEINTLPATV-----KDKAKADVNAEKE--QKNLQI 2727
QY 2501 EVOEELKRDIKIKKEKDLMOQAQATAVAPCPVTVLPAPRAPPPSPPPPVQVHTGL 2560
Db 2728 NSNDEATEEKLVAASDLNHNVEVTINQAIEDADPTNOVNE-----KXKGIG 2774
QY 2561 STPLPLVPAOSKRRKEEKSSSKKKKKMISTSEKTKDPT 2602
Db 2775 TIRDIQPLVVKKPTAKSKIESAVEKKKTETINQATNATHEVR 2816

```

```

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

Query Match 2.58; Score 363.5; DB 4; Length 10182;
 Best Local Similarity 17.3%; Pred. No. 4e-12;
 Matches 573; Conservative 488; Mismatches 1236; Indels 1019; Gaps 142;

```

QY 136 VSOQCTLMAMHVVLLKAVLRBEDTSNTTFGPADLSDVNSTLYFIDGMTPEVLAR-VY 194
Db 3117 IENOYNTAKNEAHNVL-----ENTNPTVNAVEDALRKINA-----IQPEVTKAIN 3161
QY 195 CESDKREYHNLVPOEADDPYGVENKTKYLOFLVDFLTNTIAREELMSEGYIQDDHC 254
Db 3162 ILQDKEDNSL-----VRAREK-----LDQAINQSPSLNMGTOESINNTYTKR 3204
QY 255 RVCHKLDLCCETCSAVNHECVKPPLEVEYPEDEMOCYCV-----AHKVPGVYDVC 307
Db 3205 REAQNINS-----SADTIIINGDASIEQITEKRIVEEATNALNAKQHLADPTSLK 3257
QY 308 AEIQKNPPIRHEBIGDRSRKTYWFLNRLIIEEDTENENEKKIYVSTKYVLAELIDC 367
Db 3258 TEVAK-----LSRR-----GDTNNKKPSSVSAVN-----NTIHS 3286
QY 368 LDKDYWEAE-----LCKILEMRREETHRHMDITEDLTN-----400
Db 3287 LOSETIQTEHNANTLIINKPIRSVEVNNALHEVQNLQRLDTINLLQPLANRESLKEAR 3346
QY 401 -----KAGSKNSFLA-----AANEI---LESIR 422
Db 3347 NRLESKINETVOTDGMTQOSVENVYKQAKIRQANSSIAQTLINNGDSDDEVSTETIKLN 3406
QY 423 AKKGIDINV-----KSPETEKKDKNETENDSKDAKNEEFEDQSLER---DSDDKTP 472
Db 3407 OKLSELTNINHLTVNKEPLETAKNOLOANIDQKPSIDGMTQOSVQSYERKIQEAKKIN 3466

```


Db 1157 EAAQSGSKSCQTRQTSATSTMTVMAATGAPCSAGPLIGPSMAREPGGRSPAFVQLAPLSS 1216
QY 2005 QVLTMAQLTQLTQGHGNGGLVTVIQQGQOTGOLQLP-----OG-----VTVL 2051
Db 1217 KVALSSPSIKDLPAGRHSHAVSTAAMTRSSVSGEPERMAVCSLQGGSTTVYTALE 1276
QY 2052 G---PCQQLMQ-----AAMPN-GTVQREFLPLATTAATTAATTTTTS 2090
Db 1277 ALLCPATVTVQVCSNPPCETHETGTTNTATTSNAGSAQVRCNSNPCETHETGTTTATTA 1336
QY 2091 TTACGGEOROSKLSL---QMOVHODKTLPRAOSSVG-----PAKAQPT-----AQ 2135
Db 1337 TSNAGTGQPEGGQOPAPGRPCETHHTTSTGTMSVSGALLPDATSSHRTVESGLEVAAA 1396
QY 2136 PSARPOPTOPSP-----ADPEVOTOP--EVOTOTTVSSHVPSEAPOTHAOSKPOV 2186
Db 1397 PSTTPQAGTALLAPFTQVRCNSNPCETHETGTTTATITVTSMSNSQDPPRPAASDQEV 1456
QY 2187 AAOQSPQVNOGQSPVRCVPSQ---TRIRPSTP---SGLSPGOQSOYOTTSSOPIT 2238
Db 1457 ESTQDGVNITSSSALTITTVSTLTFRATVTVTQSTFVPCSPVPPELQVSPGPROQLP- 1515
QY 2239 QPHTSLQISQGGPOQPOQVOSTOTLSSQTLNOVSVSPSRPQ-----LQIQ 2288
Db 1516 -PQQLQASSTALMGSAEVLASQTPPELPAVAVDLSSTGEPSSGQESAGSAVAATVAVOP 1574
QY 2289 POPVIAVAPQLQOQVQLSQIOSOV-----AQIQAGSGVPOQIK 2329
Db 1575 PPPTQSEVQQLSLPQELMAEAQAGTTTMTVGTLPPELAVTAALAAQAAATEBAQALA 1634
QY 2330 LQLPIDIQSSAVQTHQIQNVTVQAASVQEQQLQVQQLFDQOKKQKQOQIEIKREHTLQ 2389
Db 1635 IQAVLQAAQAAVAGTGEPMDSQAATVTVQAEIGHL-SAEGQEQATITPIVLTQOELAA 1693
QY 2390 ASMOSELIQOVVMKHNVAIEHLKOKKSMTPAE 2422
Db 1694 LVQOQQLQEAQAOQH-----HHLPTLEALPAD 1721

RESULT 6
US-08-393-703-5
Sequence 5, Application US/08393703
Patent No. 5585239
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
APPLICANT: Heirt, Winship
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-393-703-5

Query Match 2.5%; Score 353; DB 1; Length 2035;
Best Local Similarity 21.4%; Pred. No. 1.7e-12;
Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

QY 1368 TKESDSNQTTT-PSASCPESNSVNOVEDMEIETSEVKVY-----SSPTSEESNLN 1420
Db 619 TSVSSATNTSTRITITVHKSGIVTVAAQAAOVITTVGGVTKITLVKSPISVGGSL-- 676
QY 1421 DFLDENGLPINKNNENSGSKRTVITEVTTMTSTVATESKT-----VIKE-----KGDK 1471
Db 677 -----ISNLGKVMSSVQTKPVQTSAVTGAQSTGPVNIQTKGLPAGTI 721
QY 1472 QTVVSTENCAKSTVTTTTTTLKSTPSGSDIISVKEQSKTVTTVTDS----- 1525
Db 722 LKLVTSAD--GKPTTITITQASGAGTKPT--ILGISVSPSTTKRBTITTIIRTPMSAI 777
QY 1526 LTTTGGTLVTSMTVSKREYTRDKVTKLMKFSRPKTTSCTALPSV-----RKFTYKSTKS 1580
Db 778 ITQAGATGVTSSPGIKSPIITITTKVWTSCTGAPAKITIAVPRKIAIANGHQGGQVTOVYKLG 837
QY 1581 IFVLPNDLKLKARKGIREVPEYFNNA-KPALDIWYPSRPPTFGITWYRLQYKSLA 1639
Db 838 APQPGTILFTVP-MGVRLVLPVTVSAVKAVATTLV--KGTGVY--TLGTVTGV 890
QY 1640 GVSMLRLMASLRWDDMAKVPGGSGTSETSEETITTEILIKRDVCPYGLIREYCI 1699
Db 891 STSL-----AGAGHSTYSASLAPITLGTATLS----- 920
QY 1700 RKIIICPVPEPKETPTQKGLRSALRPKRPERPKO-----TGPRIETWVAEDEL 1755
Db 921 SOYINPITAVSAQITTLTAGGLTPTTMOVPSOPTOYTLITANSGV----- 970
QY 1756 WEIRAPAEVKEKAQAVEQAKRLEQOKPTVIATSTSPSTSTISPAOKVAVAP1 1815
Db 971 -----AQPV-----HDLPSILASPTTEQPLATVITADSGQGVDP- 1006
QY 1816 SGSVT-----TGKMYLTKV-----GSPATVTPQONKNFHOTFAVWKO--G 1856
Db 1007 -GTVTLVCSNPPCETHETGTTNTATTVVANLGHQPTQVOVCDROEAAASLVSTVG 1065
QY 1857 QSNQSVNOGQKVLGIIPSTGSGQTFSTFQPTATVITRPMTSGSGTSSQVITGP 1916
Db 1066 QQNGSVAVRCNSP---PCT---HETGTTNTATTAN---SNMAGOHGGSN----- 1106
QY 1917 QIRPGMTVIRTPLOQSTLGAIRTPVMVQPGAPQ--VMTQILIRGOPYSTAVS-APNTV 1973
Db 1107 -----PCETHETGTTNTATTAMSSVGANHQDARACAGIPNIRISVATGAL 1156
QY 1974 SSTPGQKS-----LTSATSTSN-----OSSASQPP-----RPQG 2004
Db 1157 EAAQSGSKSCQTRQTSATSTMTVMAATGAPCSAGPLIGPSMAREPGGRSPAFVQLAPLSS 1216
QY 2005 QVLTMAQLTQLTQGHGNGGLVTVIQQGQOTGOLQLP-----OG-----VTVL 2051
Db 1217 KVALSSPSIKDLPAGRHSHAVSTAAMTRSSVSGEPERMAVCSLQGGSTTVYTALE 1276
QY 2052 G---PCQQLMQ-----AAMPN-GTVQREFLPLATTAATTAATTTTTS 2090
Db 1277 ALLCPATVTVQVCSNPPCETHETGTTNTATTSNAGSAQVRCNSNPCETHETGTTTATTA 1336
QY 2091 TTACGGEOROSKLSL---QMOVHODKTLPRAOSSVG-----PAKAQPT-----AQ 2135
Db 1337 TSNAGTGQPEGGQOPAPGRPCETHHTTSTGTMSVSGALLPDATSSHRTVESGLEVAAA 1396

QY 2136 PSARPQOTOPSP-----ADPEVOTOP--EVQOTTVSSHVPSPAOTTHAOSKPOV 2186
 Db 1397 PSVTPQAGTALLAPFPYQVRCVSNPCETHETGTHATTVTSNMSSNODPPAPASDQGEV 1456
 QY 2187 AASQSPQSNVQOGSPVAVQSPSO-----TRIRSTP---SOLSPGOOSQVQOTTSPQIP1 2238
 Db 1457 ESTQGDVSNITSSSAITTVTSSTLTAVTVTVOSTPVPBGSVPPPELQVSPGRQOLP- 1515
 QY 2239 QPHTSIQIPSGQPOQPOVQOVSSTQTLISGQTLINQVSVSSPSRPQ-----LQIQQ 2288
 Db 1516 -PROLLQASASTALMGSAEVLASQTELPAAVDLSSTGEPSSGQESAGSAAVAVTVVQV 1574
 QY 2289 POPQVIAVPOLQOQOVVLSQIQSOV-----AQIQAOQSGVPOQIK 2329
 Db 1575 PPPQSEVVDLSLPQELMAEPAQAGTTTLVNTGLTPRELAVTAAEAAPAAATEEAQALA 1634
 QY 2330 LQLEPIQIQSSAVOTHOIQVNVTVQVQASVOEQLQVQOLRDQOQKKQOQIEIKREHTLQ 2389
 Db 1635 IQAVLIQAQQAQVAVMGTEGPMQTSAAATVTOAELGHL-SARGQEGQATTPIDIVLQOELAA 1693
 QY 2390 ASNOSEIIQKQVYKKNVAVIEHLKQKSKMTPAE 2422
 Db 1694 LVQOQOLQEAQAOQOH-----HHLPTREALAPAD 1721

RESULT 7

PCT-US93-11721-5
 ; Sequence 5, Application PC/TUS9311721
 ; GENERAL INFORMATION:
 ; APPLICANT: Lamarco, Kelly
 ; APPLICANT: Wilson, Angus
 ; APPLICANT: Herr, Winship
 ; TITLE OF INVENTION: A NOVEL EXARXOTIC TRANSCRIPTION PROTEIN:
 ; TITLE OF INVENTION: HOST CELL FACTOR
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/11721
 ; FILING DATE: 03-DEC-1993
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: PP-57503-1/RAO
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2035 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; PCT-US93-11721-5

Query Match 2.5%; Score 353; DB 5; Length 2035;
 Best Local Similarity 21.4%; Pred. NO. 1.7e-12;
 Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

QY 1368 TKESDSTQTTT-PSACPESSNVQVEDMEIETSEVKKVY-----SPTTSEESNLSN 1420
 Db 619 TSVSSATNSTRPIITVHNSGIVTAQQAQVYTVVGVGKTKITLTKSLIPSGSAL-- 676
 QY 1421 DFIDENGLPINKNEVNGESKRRTVITEVTTMTSTVATESKT-----VIKVE----KGK 1471
 Db 677 -----ISNLGKVMASVQTKPVQTSATVGAASVGPVYQIITQTKPPLPAGTI 721
 QY 1472 QTVVSTENCAKSTVTTTITVTKLSTPSGSGVDLIISYKQSGVTVTTVTS----- 1525
 Db 722 LKLVTSAD--GKPTTITTTTQASAGAKTPT--ILGISSVSPSTTKGTTITITIMSAI 777
 QY 1526 LTTTGGTLVTSMTVSKREYSTRDKVKLTKPSRPKTSAGALPSY-----RKEVTKSTKS 1580
 Db 778 ITQAGATGVTSSPGKISPIITITTKVTSQSGAPAKIITAVPKIANGHGOQGVTVLVLG 837
 QY 1581 IFVLPNDDLKLLARKGIRVPIFNPNNA-KPALDIMPSPRPTFGITRMYRLQTYKSLA 1639
 Db 838 APQOPGTLITKTPV-MGVRLVTPVTVAVKPAVTTLVV--KGTGTGT--TLGVTGTV 890
 QY 1640 GVSIMLRLLMASLRWMDMAKVPBGGSPTRETSETEITTEIIRKRDVGPVIREYCI 1699
 Db 891 STSL-----AGAGHSTASLAPITTLGIIATLS----- 920
 QY 1700 RKIICPIGVPEPKETPTPOKGLRSSALPKRPETPKQ---TGPIIETWVAEELEI 1755
 Db 921 SQVINPTAIVTSAQAQTTLTAAGGLTPTIMQVSPQTVTLITARSVE----- 970
 QY 1756 WEIRAFARERKEKAQAVEQAKKRLQOKPYIAISTSPSTSTISIPACKVAVAP1 1815
 Db 971 -----ADPV-----HDLPSVILASPTTEOPPAVTVTLADSGQVQV- 1006
 QY 1816 SGVLT-----TGKVLITTKV-----GSPATVTFQONKNEHOTFATVQK--G 1856
 Db 1007 -GVITLVCSNPPCETHETGTTNTATTIVANLGHQPQVQVCCRQEAASLVIYSTVG 1065
 QY 1857 QNSNGVVOQKVLGIIIPSTGTSQOTFSFQPRATVITRPTSSGSGTTSQVITGP 1916
 Db 1066 IQNGSVYRVCSNP---PCEFT---HETGTTNTATTAT---SNMAQOHGCSN----- 1106
 QY 1917 QIRQGMVITRPIQOSTLGAIIITRPVWQVQAPQO--VMTQIIRQOPVSTAVS-APNRY 1973
 Db 1107 -----PCETHETGTTNTATTAMSSVGNHQDARACAGTPAVIRISVATGAL 1156
 QY 1974 SSTPGOKS-----LTSATSTSNL-----QSSASQPP-----PPQG 2004
 Db 1157 EAAQGSKSQOQRTQTSATSTMTVMATGAPCSAGPLGPMAREPGRSAPVQALPLSS 1216
 QY 2005 QVKLTMAQLTQLOIGHGNOGLVIVIOGOCOTTGQOLLP-----QV-----VTVLP 2051
 Db 1217 KVLRSFSPSINDLPAGRSHAVSTAAATMRSSVAGGEPMAVCESLQGSPTTVYVTALE 1276
 QY 2052 G---PGQQLMO-----AAMPN-GTVQRFLEPLIPLATTATTASTTTTYS 2090
 Db 1277 ALLCPSATVTVQVCSNPPCETHETGTTTATTSNAGSAQRKCSNPPCETHETGTHATT 1336
 QY 2091 TTAAGTEGROSKLSP---OMOVHODKTLPPAOSSVG-----PAKAQPT-----AO 2135
 Db 1337 TSNGTGTQPEGGQQPPAGRCCEHTQTSTGTVMSVSGALLPATSSHRTVESGLEVAANA 1396
 QY 2136 PSARPQOTOPSP-----ADPEVOTOP--EVQOTTVSSHVPSPAOTTHAOSKPOV 2186
 Db 1397 PSVTPQAGTALLAPFPYQVRCVSNPCETHETGTHATTVTSNMSSNODPPAPASDQGEV 1456
 QY 2187 AASQSPQSNVQOGSPVAVQSPSO-----TRIRSTP---SOLSPGOOSQVQOTTSPQIP1 2238
 Db 1457 ESTQGDVSNITSSSAITTVTSSTLTAVTVTVOSTPVPBGSVPPPELQVSPGRQOLP- 1515
 QY 2239 QPHTSIQIPSGQPOQPOVQOVSSTQTLISGQTLINQVSVSSPSRPQ-----LQIQQ 2288
 Db 1516 -PROLLQASASTALMGSAEVLASQTELPAAVDLSSTGEPSSGQESAGSAAVAVTVVQV 1574
 QY 2289 POPQVIAVPOLQOQOVVLSQIQSOV-----AQIQAOQSGVPOQIK 2329

Db 1575 PRPTOSEVVDLSPQELMEAOAGTTTLAVTGLTPEELAVTAAEAQAQAATEEDQALAL 1634
Qy 2330 LQLPPIQOOSAVOTHOQINVTVOAASVOEOLQRYOQLRDQOQKROQOIEIKREHTIQ 2389
Db 1635 IQAVLQAQAQAVGTGEPMDSATVTOGAELGHL-SAGQSGQATTPITIVLQOELAA 1693
Qy 2390 ASQSEITQKQVVMKHNVAIEHLKOKKSMTPAE 2422
Db 1694 LVQOQOLOEAQAQOQH-----HHLPTEALAPAD 1721

RESULT 8
US-08-821-355A-7
Sequence 7, Application US/08821355A
Patent No. 5851775
GENERAL INFORMATION:
APPLICANT: Barker, Nick
APPLICANT: Clevers, Hans
APPLICANT: Korinek, Vladimir
APPLICANT: Morin, Patrice
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
APPLICANT: Sparks, Andrew
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
TITLE OF INVENTION: Interact to Prevent Cancer
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,355A
FILING DATE: 20-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UF
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5851775e
US-08-821-355A-7

Query Match 2.4%; Score 338.5; DB 2; Length 2973;
Best Local Similarity 17.1%; Pred. No. 2.2e-11;
Matches 431; Conservative 337; Mismatches 904; Indels 845; Gaps 103;

Qy 387 EIHHRMDITEDLTNRKAGSNKSFLLA-ANEELLESIRAKGIDIVNKSPEETEDKRN--- 442
Db 802 DYNRRDDNRSDFNMT---GNMTVLSPYLNTYVLPSSSSSRGSLDSRS---EKDRLSLR 854
Qy 443 -----ETENDSKAEKNREFEFS-----LEKOSDDKTYDDDEPGCK----- 480

Db 855 ERGIGLGNYPATENPETSCKRGLQISTTAAQIAKWEVYSAIHITSOEDRSSGTTIELHC 914
Qy 481 -----SEVGDf-KSEKSN-----GELSESPGAG 502
Db 915 VTDERNALRRSSAAHHTSNNTYNTKSSNRCTSCMPYAKLEYRRSSDLNSVSSSDCYG 974
Qy 503 KGASGSTRITITRLRNPDSKLSQLKS--QVVAHAHEANKI-EKEGEVLYVN----- 551
Db 975 KRCQMPKSIESYSEDESKFCSYGQYPADLAHKIHSANHHDDGDELDTPIYNSLKYSD 1034
Qy 552 --SOGELSRISTKKEVIMKGNINNYFLGQEGKYRYVHNQYSTNSFLNKHQREDHK- 608
Db 1035 QLNSGROSPSONERMAKPKHIIEDIKOSBORR--NOSTT--YPVYESTDDHLKRF 1089
Qy 609 RRHLAHKFLTPAGEFKWNGSVHGSKVLITSLRLTITTOLENNIPSSFLHPNNAHRAW 668
Db 1090 QPHFGQOECVSPYRSGANS-----ETNRVGS-----NHQINQ 1123
Qy 669 IKAVMCSKPREPALALALIECAVKPVYMLPIREFLGHTRLHMTSIEREEKYKKKE 728
Db 1124 NVQSILQOED-----DYEDDKPTVSE 1145
Qy 729 KQDEEETMQOATWVKYTFPPKHQWKQKGEYRVNGYGSMSIKTHYRFPKLPBNT 768
Db 1146 RYSEEE-----QHDEERPTVY----- 1162
Qy 789 NVNRYKSLGCTKNMDENMDES-----DKRCSRPKKIKIIEPSEKD 831
Db 1163 STYNEE---KHHVDPIDYSLKYATDIPSSOKSFPSKSSGGSOKTEHMSSEENT 1218
Qy 832 EVKSDAARGADONEMDISKITEKDDVKELDSDS-----DKP-CKEEPM 877
Db 1219 STPSSNAKR---QNLHPSSAQSRGQPOKAATCKVYSINQETITQYCVEDPTICSRCS 1275
Qy 878 EYVDDMKTESHVNCQESSQ-----VDVYVNSEGHRLT-----STKKTSS 919
Db 1276 SLSSLSAEDICNOCTOEDASANTLOIAIEIKITGRSAEDPVSEPAVSQHPRTKSS 1335
Qy 920 KLDGLLEIRIKOFTLEKORLEKIGIKIGIKSTNSKN-----LSESVI-TKAK 973
Db 1336 RLOG-----SSLSESAHKAVERFSSGAKSPSKGAQTPKPEHYVOTETLMSRGT 1388
Qy 974 EGCQSDSMRQESPMANNDQPEDLLQCCSQSDSVLRMSDPSTHTKLYPKRDVLDVSI 1033
Db 1389 SVSLDSFESRSIASSVQSP-----CSGMVSGIISPDLPSDLPQOTMPSR----- 1435
Qy 1034 RSPETKCPKONSLENDIEEVSDLASRGQEPYTSKTKGNDFFIDDSKLASAD----- 1085
Db 1436 --SKTPPEPPTAQTRKRE-----VPRKNKAPTAKRESG-----PKOAVNAVORVOY 1481
Qy 1086 --DIGTLI-----CKNK-----KPLQIEESPTIV-----SSKSALHSSV 1118
Db 1482 LPPADTLLHFASTPDGSCSSLSALSIDEPTIKDELRIIMPVQENDNENETESQ 1541
Qy 1119 PKSTNDRATPLSRAMDPEGLGCDSESNSTLENSDVTYSIODSEEDMIVONSISISE 1178
Db 1542 PKSENEMQ-----EKEAETIDSEKD-----LDDSDDD----- 1570
Qy 1179 QFRTQDQVLEVLEPLKCELVSGESTGNCEDRLPYKTEANGKPSQOKLEEPVVKCSD 1238
Db 1571 -----DIEILE--EC-IISA-----MPTKSSR-KAKKPAOTASKLPPVARKPS 1610
Qy 1239 QI---KLKNTOKKNNESESEKKGORTSPFOINGDNPKPIYLKQECLEISESVVSG 1295
Db 1611 QLPVYVLLLSQNLQPOKXVSPFG-----DDMPRY--CV-----EGTPIWF 1651
Qy 1296 NVEPKVNNINKIIPENDIKSLTVKESAIRPFINGDIYMDFNERNSEFTKSHLLSSDAE 1355
Db 1652 STATSLSDLTIESPPNEL-----AAGEGVGAQOQSEFEKRDITPIEGR--STDEAQ 1701
Qy 1356 GATRDSLETLPSTIKES-----DSTQTTTPASCP 1384
Db 1702 GGTSSV-TIPELDNKAKEGDIACINSAMPKGSKHPFRVYKIMDOVOQAASASSAP 1760

Db	802	DTNRHDDRSDNFNT	---	GNMVTLSPLANTY	LLPSSSSSSGSLDSSR	---	EKRBSLER	854				
Oy	443	-----	ETENDSKDAEKNREEFEDS	-----	LEKDSDDKTPDDDEOGK	-----		480				
Db	855	ERGIGLGNYPHATENP	PTSSRKGIQISTT	AAQIAKWEVSA	LTHISOEDRS	SGSTTELC	914					
Oy	481	-----	SEVGF	-KSEKS-	-----	GELTSPGAG	502					
Db	915	VTDERMALRRSAAH	THSNNTNYNTFK	SENSKRTCSMPYAKLETK	RSSNDLSN	VSSSDYG	974					
Oy	503	KGASGSTRITITRL	NPDSKLSQLKS	--QVVAHAHBANKL	-FKEGEVLVN	-----	551					
Db	552	--SOGELSR	LSTKREYIMKGNINNY	FTLGECKRYRYHNOY	STNSALAKKHQ	REDHDK	608					
Oy	1035	QLNSGRSPSONE	MAPKHI	IDELIKOSORSR	--NOSTI	-VPVTESTD	KHLKF	1089				
Db	609	RRHLAKHFC	LTPAGEFWMGNS	VHSGKYLITSTRLT	ITOLENNIP	SPFLPYMWASHRAM	668					
Oy	1090	QPHRGQOECV	SPYNSRANS	-----	ENRNGS	-----	NHGNO	1122				
Db	669	IKAVOMCSKPRE	ALALALECAV	PVYMLPIR	REFLGHTRLHMT	SIEREKYVKKKE	728					
Oy	1124	NVSOSLCOED	-----					1145				
Db	729	KROEBEETMO	ATWVKTVPYKHQ	VMKQKEEYRYM	GYGGMWSIK	THYREVPLPONT	788					
Db	1146	KYSEB	-----	QHEEBERTNY	-----			1166				
Oy	789	NVNTRK	SLEGTKNMDEMDES	-----	DKRCSRPKKIKIE	PDSEKD	831					
Db	1163	SIKYNEE	---KRVHDO	PIRLSKYATDIP	SSOKOSFSEK	SSSGSSKTEHMSSEMT	1218					
Oy	832	EYKGSDAK	GADGADONEM	ISITTEKODV	ELLDSDS	-----	DKP	877				
Db	1219	STPSSNAKR	--QOHLPP	SAHQSGPOKAAT	CKVSSINOET	TIOTYCVED	PIPCFSCRS	1275				
Oy	878	EVDDMKTESH	VNCOESSQ	-----	VDVNVNSEGH	LRT	-----	SYKKRTKSS	919			
Db	1276	SLSSLSA	DEIGCNOJT	OEADSANTLOIA	IEKIKIGT	HSADPEY	EVPAVAGQH	RTASS	1335			
Oy	920	KLDGLERIK	QFLPERKOLERIK	IEGKIGKIGK	STSSKN	-----	LSESPVI	TKAK	973			
Db	1336	RLQG	-----	SSLSSSA	RRKKA	VEPSSGAKSP	SGAOTAPS	SPREHYQETPLMF	SRTC	1386		
Oy	974	ECCGSDSM	ROEOPSNANND	PEDLQCCSD	SDSPVLRMS	PSHTTK	ELVPCRVLD	SVI	1033			
Db	1389	SVSSLDEFE	SNASSVOSER	-----	CSGMSGII	SPSDDL	PDSPQOTMP	PSR	1433			
Oy	1034	RSPETKCP	KONSIE	NDIEEYVSLAS	RGOEPTKSG	KNGDF	ITDSK	LASAD	-----	1085		
Db	1436	--STPP	PPPTQATKRE	-----	VPKNAKAPT	LEKRESG	-----	PKQAAVNA	AVOROV	1483		
Oy	1086	--DIGTLI	-----	CKNK	-----	KRLIOBED	TIY	-----	SSSKALHSY	1118		
Db	1482	LPDADTL	LHFATEST	PDGFCSS	SLSLASL	DEPFIQ	KDVLIRMP	VOENDG	NGNTESEQ	1541		
Oy	1119	PKSTNRD	ATPLSRAM	DEEKLGCDS	ESNSTLNS	SDTVI	ODSS	SEEDM	IVONS	MEISIE	1178	
Db	1542	PKSENENQ	-----	EKA	EKTI	DSEKD	-----	LDDSDDD	-----	1576		
Oy	1179	QFRTRED	VDVLEPLK	ELVSGEST	GNCDRL	PVYKTE	ANGKPP	SOQK	LEER	PVYNKSD	1238	
Db	1571	-----	DIELLE	--EC	-IISA	-----	MPTKSSR	-KANKPA	OTASK	LEPPYAKRPS	1610	
Oy	1239	QI	---KLKNT	DDKKNNR	RESEKKG	ORTSTFOJ	NGDNPK	ITYLGE	CLKE	ISESRVYG	1295	
Db	1611	QLPYVYK	LLPSONR	LOPKHYS	FTPG	-----	DMAPRY	---	CY	-----	EGTPINF	1651
Oy	1296	NVEPKVNN	IKIIP	ENDIKSL	YVKE	SARIP	ENGDI	MEDE	FNERS	SETKSH	LLSSDAE	1355

Dh	1652	STATSLDOLITESPNNEL	-----AAGCGVGAAGSGEFEKDRITPEGR	-STDEAQ	1701
Qy	1336	GNYRDSLETLPSTES	-----	-DSTQTTTPSACP	1384
Dh	1702	GKRTSV-TIPELDONDAEBGDI	LAECINSAMEKGSNHPFRVKKIMDVOVQASASSAP	-----	1760
Qy	1385	ESNSNVOYDEIETSEVKKYTS	-----	-SPITSEESNLSN	1420
Dh	1761	NKNQLODKR	--KRTSPYKPLPONTETRTYRVKNMADSKNNLNAERYFSDNKSQKNLKN	-----	1818
Qy	1421	DFIDENGLPIRKNENVGE	-----	-----SKR	1443
Dh	1819	NSKDPENDLPINNEDRVGSPAFSD	PHHYTPTEGTPCYGSRNDSLSELDDEDDVDLSREK	-----	1878
Qy	1444	TWTEVTMTSTVATSESKYIAK	EKGKQVVSSTENACKSYT	-----TTTTV	1493
Dh	1879	---AELRAKENKESEKVTSH	TELSNMOQSAKTAQIAKOPINRGPKPILOKOSTPQ	-----	1935
Qy	1494	TKLSTPSTGSGVDITSVKEOSK	TVTTVTYVDSLTTTGATLVMTVSKESTYRDXVKLAK	-----	1533
Dh	1936	SSKOIPDGGATD	--EKLONPAIENTPVCFSS	---HNSLSLSDIDQDNKKKEPILKE	1989
Qy	1554	FSRP-----KTRSGTALPSTR	-----FYTKTKKSIFYLPNDL	-----	1588
Dh	1990	TEPPSOGEPSKPOASGYAPKSF	HEVEDPYVCFSRNLSLSLSDSEDDLOJCISAMPK	-----	2049
Qy	1590	-KTLARKGIREVYFVYNNKP	-----ALDIWYPSRPTEFGT	-----WRY	1630
Dh	2050	KKKPSRLKGDNE	-----KXSPRMGGLGEDITL	KIQRDPSRHGSLSPSEMPDKA	2103
Qy	1631	RLQTVKSLAGVSLMLRLMLAS	SLRMDMAKVPBGSGSTRETSETETTELIKRDRVCP	-----	1690
Dh	2104	IOEGANSIV	-----SLIHQAAAACL	-----SRQASDSD	2140
Qy	1691	YGIREVYCIKRIIOPICVPE	PKETTPPOKGLRSSALBPKRPKPPKQGVIIETWAE	-----	1750
Dh	2141	-GI-----SLGSPHLTPDQEK	PTSMKGR--TLKPEKST	-----LETKIE	2182
Qy	1751	EE-----LELMRAFAERVENEK	QAQAEQAKRLEQOKP	-----TVAT	1793
Dh	2183	SESKIGKCKVYKSLITGVR	NSSE-----ISGQMOPLOANMPSISRGTHI	-----	2233
Qy	1792	STSTSTSTSTISPAOKVW	AVAPISGSYTTGTAMULTTVKGSBAVTTFOQKNFHOTAT	-----	1851
Dh	2234	PGVRNSSSTSPVSKCPPLK	TAPASKSPSGQATTPSPGAKPV	-----	2278
Qy	1852	VWKOOSMSGVVOQKVLG	LIPSNTCSOQPTSPQPRATVYTRPNTSGSGGCTSSQ	-----	1911
Dh	2279	-----KSLSLVAQTOIG	-----GSKR	-----AFSNGSRDSTPS	2310
Qy	1912	VITGPQIRPGMTVTRPLQ	SLGKALITRPVWOBG	-----APQVMQITRGQPVSTAV	1967
Dh	2311	-----RPAQOPLSRLOSPGR	NS-----ISGRNGISPPKMLSQLPRTSSPSTA	-----	2354
Qy	1968	SAPNTVSTSPQOKLTS	-----ATISNIOSSASOPPRQOQVKYLMALQTLQTL	-----	2018
Dh	2355	-----STKSSGSKMSYTP	SGMSQOQLTKOTGLSKNASSIPSESSAS	-----KGLNOMN	2406
Qy	2019	GHGNGOGLTVVIOQOGOT	TOLO-----LIPQGVTVIPGGOOLMQAMNCTVQREFL	-----	2071
Dh	2407	GNGANKKVELSRMSST	SSGSESDRSRPVLVQSTFIEARISPLTRKLESSAS	-----	2461
Qy	2072	FTPLATTAATTA	STTTVTYVSTTAAGTGEOROKSLSP	-----QMOVHOD	2120
Dh	2462	FESLSPSSRPASPIR	SOAQTPV-----LSPILPMJSLTSHSVQAGGWRKLPNUL	-----	2511
Qy	2121	SSSVGPAKAOPTOAGSPARO	POPTOPOSPAPQPEVQOTQYTVSHVPSSEAOPTHAQ	-----	2180
Dh	2512	SPITIEYNDGRAKH	DLARSHSESPRLPINSRGTYKRE	-----HSHKSSLPLRVSTWMTKRG	2568
Qy	2181	SSKPOVAAQSQPOS	NVQSGSPRVVQSPQTRIRPSTQSL	-----PGQOSOVQTTTSQPIPIQ	2239
Dh	2569	SSSLSLSSSSSESKAKSE	DEDEKHNVSJSG--KQSKENOVSSKGTWYRKIKENEFSP	-----	2622


```

OY 1554 FSRP-----KTRSGTALPSTRK-----FVTKSTKSI EYLPNDL----- 1589
DB 1990 TEPPDQGEPSKPOASGYAPKSFHVEDTVCFSRNSLSLSDISEDDLQECISSAMPK 2049
OY 1590 -KLLARKGIREVPYNNYNAKP-----ALDIMPSPRPFTGTT-----WRY 1630
DB 2050 KKPSPRLKGDNE-----KHSPPNMGILGEDTLDDKIDQRPDSEGLSPDENFDWKA 2103
OY 1631 RLQTVKSLAGVSLMLRLMLASLMDMAKVPFGSGSTRETETSETETTTETITIKRDVGP 1690
DB 2104 IQGANSIV-----SLHQAAAACL-----SRASSDSD-----SILSLKS--- 2140
OY 1691 YGIRFEYCIRKIICPIGVETPEKRETPPORKGLRSSALRPKRETPKQNGPVIIETWAE 1750
DB 2141 -GI-----SLGSPHLLTPDQEEKPEFTSNKGBR--ILKGEKST-----LETKKIE 2182
OY 1751 EE-----LEMEIRAFERVENEKRAVEQQAOKKRLBOOKP-----TVIAT 1791
DB 2183 SESKIGIKGKKVYKSLITGKVRNSSE-----ISGOKKQPLQANMPSISRGRTMHI 2233
OY 1792 STTSPTSTSTTSPAKQVAVAPISGVTTGTAKVLTTKVGSPATVTFQONKNFHQTFAT 1851
DB 2234 PGVRNSSSTSPYKSKGPPLKTPASKSPSEGOATSPRKAQSV----- 2278
OY 1852 WYKQGSNSGVVOVGKVLGIIPSSGTGTSQOFTTSFOPTATVYTRPNTSGSGGTTNSQ 1911
DB 2279 -----KSELSPVARQTSQIG-----GSSK-----APSRGSRDSTPS-- 2310
OY 1912 VITGPQIRGMYTIRTPLOQSTIGKALIRTPVAVQV-----APOQVOTQIIRQPVSTAV 1967
DB 2311 -----RPQOPLSRPLOSPPGNS-----ISPRNGISPPNKLQSLQPRSSPPTA- 2354
OY 1968 SAPNTVSTPGQKSLTS-----ATSTNIOSSASOPRPOQGVKLTMAQLPOLQ 2018
DB 2355 -----STKSSGSGKMYSTSPGKMSQMLTKOTGLSKNASTIPRESAS-----KGLNOMN 2406
OY 2019 GHGNGOGLTVVIOGQOQTGLO-----LIPQGVTVLPQGOQLMQAAMPNGTVORFL 2071
DB 2407 GGNANKKVELSRMSTKSSGSESDRSERPVLVQSTFIKAPSPTLRKLFEFSAS----- 2461
OY 2072 FPLPLATTATTTTSTTTVSTAGTGEORSKSP-----QMVHND-----KITLPPAQ 2120
DB 2462 FESLSPPSPASTPNSOAOQPV-----LSPSLPDMSLSTHSSVQAGMRKRLPPL 2511
OY 2121 SSSVGPAAKOPOTAPASAPPOPTOPSPAPQPEVOTOPPEVOTQTVSSHVPSAQPHTAQ 2180
DB 2512 SPTEIENDGRPAKRHDIAHSHSPSLRPIINRGCTWKR-----HSKISSSLPRVSTMRRTG 2568
OY 2181 SSKPQVAAQSPQSNVQGSPPVRQSPSOTRIRPSTPSQLS--PGQOSQVOTTTTSPPIQ 2239
DB 2569 SSSSILSASESESEKAKSEDEKHVNSISGT--KOSKENQVSAKWTIKENEFSP----- 2622
OY 2240 PHTSLQIPSGQPQSPQVOVOSTOTJLSSGOTLQNVSSPSRQLOLQOPQPOVIAPOL 2299
DB 2623 -----TNSISQTVSSGAT-----NGAESKTLIYQAPAVSKTEDEV 2657
OY 2300 QOOVQVLSIQSOVVAQIOAOGSG-----VPOQIKLQDLQIOQSSAVQTHQIONV 2350
DB 2658 WVRLE-----DCPINNPRSGRSPGTGNTPRVIDSVSKANPNIKDSQDNQAK--QNV 2706
OY 2351 ----VTVOAASVQEOLOVQOQLRDQOKKQOQIEIKREHTLQASNOSEIIQK-----QV 2401
DB 2707 GNGSVPMRTVJLENRLNSFIQVDAPOKGTETIKPGQNNPVPVSETNESSIVERTPSSSS 2766
OY 2402 VMKHNHV-----IEHLKQKSMTPARENOIRIIVONQWYKTLIDITDKEEQ 2449
DB 2767 SSKHSSPSGTVARVTPFNVPNPSPRKSSADSTARSQIPTPVNNNTKKRDSQTDSTESS 2826
OY 2450 AAKKRREE-----SVEOKRSK-----QNAKTLALTFKHQKQELAEILKRALLD 2495
DB 2827 GQSPSRHSGSYLVTVSVKGRMKLRKFYVNYNCYIDILFQMK--LKEFKCFKFLLE 2881

```

```

RESULT 11
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
;
GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; City: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
;
US-07-741-940-7
;
Query Match 2.4%; Score 337; DB 1; Length 2842;
Best local Similarity 17.1%; Pred. No. 2.5e-11;
Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;
;
OY 387 EIRHMDITDLTNKARGSKSFLAA-ANPEILSIRAKGIDIDNVKSPETEKDN--- 442
DB 801 DTRHDDNDRSDNFTNT--GNMTVLSPLYLNTVLPSSSSSSGSLDSSRS---EKDRSLER 853
OY 443 -----ETENDSKDQEKNRREPEDS-----LEKSDDKTTPDDPEQCK----- 480
DB 854 ERGIGLGNHPATENPQSTSSKRGLOJSTTAQIAKAVEVSAIHTSOEDRSSGTTLHLC 913
OY 481 -----SEVGF--KSEKSN-----GELSESPGAG 502
DB 914 VTDERNALRRSSAAHHSNNTYNTFKSBNRQCSMPYAKLEYKRRSSNDLSNVSSSDQYG 973
OY 503 KGASGSTRITTRLNDPSKISQLKS--QVYAAAHAEANKL-FREGKEVLYVN----- 551

```

```
Db 974 KRGOMKPSIESHDESCFCSGYQADLAHKIHSANHMDNDGELDPTINYSKXSYDE 1033
QY 552 --SOGEISRLSTKKEVIMKGNINNYFELGEGKYRYVHNOYSTNSFALNKHQHEHDOK- 608
Db 1034 QLNQSGRQSPQNEBWARPKHIEDEIKOSQORQSR--NOSTT--YVYTESTDCKHLKE 1088
QY 609 RRLHAHFCLTPAGEFKWNSVHOSKVLITLRLITTOLENNIPSSFILPNNASHRANW 666
Db 1089 QPHGQOEVSYPYRSRQANS-----ETNKGVS-----NHGINQ 1122
QY 669 IKAQOMSKSRREFALALAECAVKPYVMPPIWEFLGHTRLHMTIEREKEKVKKE 728
Db 1123 NVSGSLQOED-----DVEDDKPTVYSE 1144
QY 729 KQOEESTMOQATWYKTFPVKHQVWKQGEYRVTGYGMSWISKTHYVRVFKPLDGT 788
Db 1145 RYSEEE-----QHEEERPTNY----- 1161
QY 789 NVNYSKSLGTRKNMDENMES-----DKRCSRPKKIKIEPDSEKD 831
Db 1162 SIKYNEE--KRVHDPIDISLKYATDIPSSQKOSFSPKSSGSSKTEHHMSSSENT 1217
QY 832 EVKGSDAKAGADQENMDISKTEKKDQVKELDSDS-----DKP-CKEPEM 877
Db 1218 STPSSNAKR--QNLHPSSAQSRSGOPKAKATCKVSSINOETIQYCVEDTPICFGRCS 1274
QY 878 EVDDDMKTESIVNCOESSQ-----VDVYVNSGFIHLT-----SYKKTKSS 919
Db 1275 SLSSLSAEDEICGNQTOEADSNANTLOIAIEIKKIGTRAEADPVEVPAYSOHPRTKSS 1334
QY 920 KLIDLERRIKQFLEKORLEKIKIGIKIGIKTSTNSKN-----LSESPVI--TKAK 973
Db 1335 RLQO-----SSLSSESAHKHAKVEFSSGAKSPKSGAQTPKSPPEHYOETPLMFRCIT 1367
QY 974 EGCSOSDMROEQSPNANNQPEDLIQCCSOSDSSVLRMSPSHTTNKLVPKRYLDVSI 1033
Db 1388 SVSLSDSFESHRSIASVQSEF-----CSQWVSGIISPSPDLPDPCQTMPPSR----- 1434
QY 1034 RSPTKCPKQNSIENDIEEKVSDLASRGOEPTKSKTKGNDFFIDSKLASAD----- 1085
Db 1435 --SKTPPPPTQOTKRE-----VPKNKAPTAEKRESG-----PKQAAVNAVQRYOV 1480
QY 1086 --DIGTLI-----CKNK-----KPLIOESDPIV-----SSSKSALHSV 1118
Db 1481 LPQADTLHLFATESTPDGFCSSSLSALSIDEPTIQDVELRIMPVQOENDNGETSEQ 1540
QY 1119 PKSTINDADAPLPSRAMPEKGLCDSESNTLENSDPTVISIQDSEEDMIVQNSNEISE 1178
Db 1541 PKESNENO-----EKEAEKTISEKD--LLDSDSD-- 1569
QY 1179 QFRTREDDVELEPLKCELVSGESTGNCEDLPPVKGTEANGKRPQOKLEERPVNKCSD 1238
Db 1570 -----DIELE--EC-IISA-----MPTKSSR-KAKRAQAQASKLPPVYAKKPS 1609
QY 1239 QI---KUKNTDKKNNRESEKKGQRTSTFQINGKDNKPKIYLKGCLEISESRVSG 1295
Db 1610 QLPYKILPSPQRLQOPKHVSFTPG-----DDMPRVY--CV-----EGTPINF 1650
QY 1296 NVEKVVNNIKIIPENDIKSLTVKESAIRPPINDVIMEDFNENNSSETSKSHLSSDAE 1355
Db 1651 STATSLDLDLTIESPPNEL-----AAGECVRGCAQSGEERKDTIPTER--STDEAQ 1700
QY 1356 GNYRDSLETPSTKES-----DSQUTTPSASCP 1384
Db 1701 GGTSSV--TIPELDNKAEBEDIIAECINAMPKGSKHPRPVKKIMDQVOQASASSAP 1759
QY 1385 ESNSVNOVEDMEIETSEYKAVTS-----SPITSEESNLSN 1420
Db 1760 NKNQLDGKK--KKPTSPVKPIPOUTERYRVRKNADSKNNLNAERVPSDNKDSKQMLKN 1817
QY 1421 DFIDENGLPIKNKNVNGE-----SKRK 1443
Db 1818 NSKDFNDKLPNNEDRVGSPAFDPHHYTPLEGTPYCFGRNDSLSLDFDDDDVLDLSREK 1877

QY 1444 TVITEVTTMTSTVATESKTVIKVEKGDQOTVSVSTENCAKSYT-----TTTTV 1493
Db 1878 ---AELKRAKENESEKAVTSHTELTSNDOOSANKTQOIAQOPINRGOPKILOKOSTPEQ 1934
QY 1494 TKLSTPSTGSDVIDISVKEOSKTVVTTVYVDSLTGTLVATSMVSKESYSTRDKVLMK 1553
Db 1935 SKSDIPDRGAATD---EKLONFALENTPVCF--NHSSLSLSLDIDQENNNENEDRIKE 1988
QY 1554 FSRP-----KTRSGTALPSYRK-----FYTKSTKSIYVLPNDL----- 1589
Db 1989 TEPPDSQGEPSKQASGVAKPSFEHEDTPVCFGRNSSLSSLSLSDSEDDLQOECISSAMPK 2048
QY 1590 -KKIARKGIGIREVYFYNNKP-----ALDIWYPPSPRPTGIT-----WRY 1630
Db 2049 KKRPSRLKQDNE-----KSPRNMGIILEDLTLDKIDQRPDSEHGLSPDSEBNDWKA 2102
QY 1631 RLQTVKSLAGVSLMLRLMASLRMDMAAKVPPGGSTRTETSETTEITTEIIRRDVGP 1690
Db 2103 IQEGANSIV-----SLHQAAAACL-----SRQASSDSD--SILSLKS-- 2139
QY 1691 YGIREYCIKKIICPIGVPEPTPKETPPQKGLRSSALRPKRPETPKQTPVLIETVWAE 1750
Db 2140 -GI-----SLGSPFHLPDQOEKPTSNKGP--ILKPGEKST-----LETKIE 2181
QY 1751 EE-----LELWEIRAFEREVEREKAQAVERQAKRLQOQK-----TVIAT 1791
Db 2182 SESEKIGIKGKKVYKSLTTCGVRSNSE-----ISGOKQPOLQANMPSISGRITVHI 2232
QY 1792 STYSPSTSTSTISPAQKVVAVPISGSVTGTGKVVLTGKVPATVYFQONKNFHQTEAT 1851
Db 2233 PGVANSSTSTSPVSKKGPPLKTPASKSPSEGQVATTSPRGAKPSV----- 2277
QY 1852 WKQOGNSGVVOYQAVLGIIPBSTGTSOOTFISFPQPRATVYIRENTSGSGGTTNSQ 1911
Db 2278 ---KSELSPVARQTSQIG-----GSSK-----AFSRGSGDSTPS-- 2309
QY 1912 VITGPQIRPGMTVYIRPLQOSTLGKALIRTPVMVOPG---APQVWTOIIRGPVSTAV 1867
Db 2310 -----RPAQOPLSRPIQSPGRNS-----ISPRNIGISPNNKLSQLPRISSPSTA- 2353
QY 1968 SAPNTVSTSPGQKSLTS-----ATSTNSIQSSASQPPRPOQOGVKTLMQLOLTQ 2018
Db 2354 ---STKSSGSGKMSYTPSGQMSQONLTQOTGLSKNASSIPRSESAS---KGLNDMNN 2405
QY 2019 GHGNOGLTVYIQGQGTGLOQ-----LIPGVTVLPQGGQQLQAAAMPNGTVQRL 2071
Db 2406 GNGANKKVELSRMSSTKSSGESDRSERPVYLRKOSTIKAPFTLRLKLEESAS----- 2460
QY 2072 FTPLATTATTATTTTIVSTTAAGTGQROSKLSP-----QMOVHOD-----KTLPPAQ 2120
Db 2461 FESLSPSSRPASPIRQAOQIPV-----LSPSLPMSLSTHSSVQAGWRKLPPNL 2510
QY 2121 SSSVYPAKAOPTAOPAPROPOTOPOSPAPQPEYOTOPPEYOTQTVYSSHVSEAOPIHAQ 2180
Db 2511 SPRTIEYNDGRAPRAKHDAIRHSHSESPRLPINRSQWTRRE---HSHKNSLSPRVSTWRTG 2367
QY 2181 SSRPOVAOQSPQSNVQOGSPVPRVQSPQTRIRPSTRSOLS--PQOQSOVOVOTTSQPIPIQ 2239
Db 2568 SSSSILSASSSESKAKASEDEKHVNSISGT--KOSKENOVSAKGTWIKIRENEFSP----- 2621
QY 2240 PHTSLQIPSOQOPQOSOPQVOSTOTLSSQOT 2270
Db 2622 -----TNSTSQTVSSGAT 2634

RESULT 12
US-08-289-548A-7
; Sequence 7, Application us/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
```

APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 2.4%; Score 337; DB 1; Length 2842;
Best Local Similarity 17.1%; Pred. No. 2.5e-11;
Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;

QY 387 EIRHMDITEDLTKNRKSGSKSFLAA-ANEELIESTIRAKKGDIDNVKSPETEKKDKN---442
DB 801 DTRRHDDNNSDNNT---GNMIVLSPLYLNTVTPSSSSSGSLDSSRS---EKDLSLER 853
QY 443 -----ETENDSKDAEKNEEFEDOS-----LEKSDDKTDPDDPEOQK-----480
DB 854 ERGIGLVNHPATENPGTSGKRGLOISTTAQOLAKYMEVSAIHITQOEDSSSTTELHC 913
QY 481 -----SEYGF-KSEKSN-----GELSESPGAG 502
DB 914 YVDERNALRRSSAAHTSHSTYNTKTSKNSNRTCSMPYAKLEYKRSSNDLSNVSSTGSGYG 973
QY 503 KGASGSTRITRLRNPDSKLSQLKS--QOYAAAHAEANKL-FKEGKEVLVYN-----551
DB 974 KRQGMKPSIESSEDESEKFCSTGYRPADIANKIHSHNHDNDGELDPPINYSLSKYSDE 1033
QY 552 --SQGEISRLSTKKEVIMKGINNRYFKLQEGKYRYVHNOYSTNSPALNKHQREDHDK- 608
DB 1034 QLNSGQSPSONRWRMRPHIIEDEIKQSPQGRS--NQSTT--VPYVTESTDDKHLKF 1088
QY 609 RRLIAKFCILTPGGEFKMGSGVSGKVLITSLRLITTOLENNIPSSFLHPNASHRANV 668

DB 1089 QPHFGQDECVSPYRSKAGANS-----ETNRVGS-----NHGINQ 1122
QY 669 IKAVOMCSKPREFALALALECAVKPVVMLPIWREFLGHTRLHMSTIEREEKYKKE 728
DB 1123 NVSOSLQOED-----DVEDKPTVYSE 1144
QY 729 KQDEEETMOQATWVKTFPVKQVWKQGEYRVTGYGWSMISTKHVYRPVPLRGNT 768
DB 1145 RYSEEE-----QHDEEERPTVY-----1161
QY 789 NVNKRKSLSECTKNMDENMDES-----DKRCSRPRKIKIEPOSEKD 831
DB 1162 SIKYNEE---KRVDPIDYSLKATDIPSSQKSEFSKSSGSSKTEHSSSENT 1217
QY 832 EVKSPDAKAGADONEMDISKITEKKDDVKELDSDS-----DKP-CKEEPM 877
DB 1218 STPSSNAKR--QNLHPSSAQSRSGQPOKAATCKVSSINQENTIQYCVBDFICFRCS 1274
QY 878 EVDDMKTESHVNCQESSQ-----VDVNVSEGFHLRT-----SYKKTKSS 919
DB 1275 SLSSLSAEDEICGNQTOQADSANTLOIAIEIKETIGTSAEDPVSEVPVSOHPRTKSS 1334
QY 920 KLDGILERIKQFTLEKQKLEKIKLEGIKIGKSTNSKN-----LSESPVI-TKAK 973
DB 1335 RLOG-----SSLSSSAHHKAAVESSGAKSPKSGAQTPKSPPEHYVOETPLMFSRCT 1387
QY 974 EGCQSDSMROEOSPNNANDPEDLIGCSQSDSSVLRMSPSHTNTKLYPKDRVLDDVSI 1033
DB 1388 SVSSLSDFESRSIASSVQSP-----CSGMVSGIISPLDPSPOGTMPPSR-----1434
QY 1034 RSEPTCKPKONSIENDIEEKVSDLASRGQPTKSKTGNDFTIDSKLASAD-----1085
DB 1435 --SKTPPPPPQATQTKRE-----VPKNKAPTAEKRESG-----PKQAVNNAVQVRVQ 1480
QY 1086 --DIGTLI-----CKNK-----KPLIOESPTIV-----SSKSALHSSV 1118
DB 1481 LPADVTLHFATESTPDGFCSSSLALSDEPFIQDVLRIMPVQENDKNETESEQ 1540
QY 1119 PKSTNDRDATPLSRAMDEEKLKCDSESNSTLENSDPTVASIODSEEDMIVQNSNEISE 1178
DB 1541 PKSENEO-----EKEAKTIDSEKD--LDDSDDD-----1569
QY 1179 QFRTREOVVEVLPKLCGLVSGESTGNCEDRLPVYKTEANGKRPQOKLEBRPVKNCSD 1238
DB 1570 -----DIEILE--EC-IISA-----MPTKSSR-KAKKPAQASKLPPVARKPS 1609
QY 1239 QI---KLNKTDKKNNENRSEKKGQRTSTFQJNGKDNKIKYLGKELKEISERVSG 1295
DB 1610 QLPVYKLLPQONRLOQKHVSFTPG-----DDMRVY---CV---EGTPINP 1650
QY 1296 NVBPKVNNINKIIPENDIKSLVTKESAIRPFINGDVIEMFENRNSSETKSHLLSSDAE 1355
DB 1651 STATSLSDLTIESPPNEL-----AAGGVGGAOSGEFEKRDITPIEGR--STDEAQ 1700
QY 1356 GNYRDSLETLPTSKES-----DSTQTTTPSASCP 1384
DB 1701 GKGTSVY-TIPELDNKAKEGDILABCINSAMPKGSKHFVYKKIMDQYQQAASASSAP 1759
QY 1385 ESNVMOVEMETSEVKKVYTS-----SPTTSEESLSN 1420
DB 1760 NKQLODKK--KKPTSPVKPIPONTETRTVRKRNADSKNNLNAERVFSNKKDKKONLKN 1817
QY 1421 DFIENGGLPIKNKNENGE-----SKRK 1443
DB 1818 NSKDFNDKLPNNEDRVRGSAFDSPIHHYPIEGTPYCFGSNDLSLIDPDDDDVDLSREK 1877
QY 1444 TVTTEVYTMSTVATESKTVIKVEKGDQTVVASTENCAKSTVY-----TTTIV 1493
DB 1878 --AELRKAKENKESAKVYSHNELSNSQOASAKTQAIAPINRQOPKPIILQKQSTPPQ 1934
QY 1494 TKLSTPSTGSDVDTISVKEOSKTVVYTVYDLSLTGGLVYTMVYSKEYSTDDKYLAK 1553
DB 1935 SSKDIPDRGAATD---EKLONFAIENTPVCFSS--HNSLSLSLSDIDQENNNKNEPIKE 1988


```

QY 1554 FSRP-----KTRSGTALPSYK-----FVTKSTKKSIFVLPNDL-----1569
D 1989 TEPDSDGERSKPOASGTAPEKSHVEDT PVCFSRNSSLSISDSEDLQECTSSAMPK 2048
QY 1590 -KTLARKGIREVYFYNMKNP-----ALDIWPPSPRPRTG1T-----WRY 1630
D 2049 KKKPSRLKGDNE-----KHSFRMWGILGDBDLTDLKDIQRPDSEHGLSPDSEFDMKA 2102
QY 1631 RLQTVKSLAGVSLMLRLMLASLRDMDMAKVPBGSGSTRTTETSTETTTETITRRDYCP 1690
D 2103 IQEGANSIV-----SLHQAAMAAACI-----SROASSDSD-----SLSLKS-----2139
QY 1691 YGIFEFYCIKRIICPIGVPEPPKPTPPORRGLRSALRKRPRPTKOTGPFVITETWAAE 1750
D 2140 -GI-----SLGSPFLUTPDQEKPTSNKGR--ILKGEKST-----LETKEIE 2181
QY 1751 EE-----LBLEIRAFARERKEKAQAVBQAKRLEQOKP-----TVIAT 1791
D 2182 SESKIGKGGKKVYKSLITGKVRNS-----ISQMKQPLQAMMPSISRGRTMHI 2232
QY 1792 STSPSTSTSTISPAKQVMAVAPISGSVTTGKAVLTTKGSPATVTFQONKNHQTAF 1851
D 2233 PGVRNSSSTSPVSKKGPPLKTPASKSPSEGTATTPRGAKPSV-----2277
QY 1852 WVKOGOSNGVVOYQKVLGIIPSTGTSOQTFSPQRTATVIRPNTSGSGGTSNSQ 1911
D 2278 -----KSELSPVARQTSQIG-----GSSK-----ASRSRSDSTPS--2309
QY 1912 VTGPPQIRPGMTVIRTPLOOSTLCKAIIFFPVWVOPG---APOQVMTQIIRGQPVSTAV 1967
D 2310 -----RPAQOPLSRPIQSPGRNS-----ISPGRNGLSPNKLSQLPRTSSPSTA-2353
QY 1968 SAPVTSSTPGOKSLTS-----ATSTNIOSSASQPRPQGGVQKLMALQTLQ 2018
D 2354 -----STKSSGSGKMSYTBEGRQMSQONLTKOTGLSKNASSIIPRSAS-----KCLNOMNN 2405
QY 2019 GHGNGOGLTVVIOGQGTGLO-----LIPOGVTVLPBGQOLMOAAMPNGTVORFL 2071
D 2406 GNGANKKVELSRMSSSTKSSGSESDRSERPVLPQSTFIKEAPSTLRKLEESAS-----2460
QY 2072 FTPLATATATSTTTTIVSTTAAGTGEOROSKLS-----OMOVHOD-----KTLPPAQ 2120
D 2461 FESISPPSRPSPRPSQAQIPV-----LSPLSDMSLSTHSSVQAGWKRLPNL 2510
QY 2121 SSSVGPAAKAPQOTAPQAPQOTOPQSPAPQEPVQOTPEVQOTOTTVSSHVSEAOPTHAQ 2180
D 2511 SPTEIYNDGRAPRAKHDIARSHSESPSKLPINRSQTKRE---HSHSSSLPRTVSTWRTG 2567
QY 2181 SSKPQVAAQSQPQSVQGSFVRVQSPQTRIRPSTPSQLS--PGQOSQOVQTTTSQPIPIQ 2239
D 2568 SSSSILSSSESESEKAKSEDEKHNYSISGT--KOSKEWQSAKGTWIKIKENEFSP----2621
QY 2240 PHTSLQIPQSQQPOQPOQVQVQSTOTLSSGOT 2270
D 2622 -----TNSTQTVSSGAT 2634

```

```

; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESS: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-452-654-7
;
; Query Match 2.4%; Score 337; DB 1; Length 2842;
; Best Local Similarity 17.1%; Pred. No. 2.5e-11;
; Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;
;
; QY 387 EIHHRMDITEDLTNKAAGSKSFLAA-ANPELLESTIRAKKGDIDNVKSPETEDKN---442
; D 801 DTNRHHDNRSDNFT--GNMTVLSPLYLNTTVLPSSSSKSLDSRS---EKDRLER 853
; QY 443 -----ETENDSKDAEKNEFEEDQS-----LEKDSDDKTPDDPEQK-----480
; D 854 ERGIGLGNHYHPTENPNTSSKRGQISTTAAQIAKWEVSAIHTQEDSSGSTEELHC 913
; QY 481 -----SEVQDF-KSEKSN-----GEISESPGAG 502
; D 914 VTDERNALRRSSAAHTSHNTYNTFTKSENSNRTCSPYAKLEYKSSNDLSNVSSSDGY 973
; QY 503 KGASGSTRITIRLKNPDKSLQKS--QVAAAHENKLFKREKRELVVN-----551
; D 974 KRQMKPSIESYSEDDESKCSYGQYADLAHKIHSANHMDNDGELDPTINSLKYSDE 1033
; QY 552 --SOGELSRISTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNFAINKHQREHDK- 608
; D 1034 QLNSGRQSPSONERKMAKHIIEDEIQSQGRS---NOSTI--YVYTESDDKILKF 1088
; QY 609 RRLAHKFCLLTPAGEFWKNGSVHSGKVLITSLRLITTDENNIPSSFLHPNASHRANW 668
; D 1089 QPHFGQEQECVSPYRSGANGS-----ETNRVGS-----NHGING 1122
; QY 669 IKAVQMSKREPALALALIECAVKPVVMLPIRREFLGHRILHRMSIEREKEKVKYKKE 728
; D 1123 NVSOSLCOED-----DYEDDKPTNYSE 1144

```

Qy	729	KQ0EEETMOAOATVWKTTPFVKHQVKKQGEERYATYCGJGMSWISKTHVYREPKLPNGT	788
Db	1145	RYSEEE-----OHEEERPTNY-----	1161
Qy	789	NVNRKRSLEGTKNMDEMOMES-----DKKCRSPKPKIKIEPDSKED	831
Db	1162	SIKINEE-----KRHVDPIDYSLKYATDIPSSOKOSFSPKSSGSSCKTEHSSSSENT	1217
Qy	832	EVKSGDAKAGADONEMDISKITEKKDODVKELLDS-----DKP-CKEEPM	877
Db	1218	STPSNMKR-----QNOJHPSSAQSRGQPOKAACTCKVSSINQETIQTYCEDPIPCFMSCS	1274
Qy	878	EVDMDKATKESHVNCQESQ-----VDVYVNSSEGHFLT-----SYKKITSS	919
Db	1275	SLSSLSAEDEICGNOTTOADANTLOIAIEKIKGTGTSAEDPYSEVPAYSQHPRRTSS	1334
Qy	920	KLDGLERRIKQFTLEKQLEKIKLEGGIKGIGKSTNSKN-----LSESPV-TRAK	973
Db	1335	RLOG-----SLSSSAHKAVERSSCAKSPKSGAOTPKSPREHYOETPLMFSTCT	1387
Qy	974	EGCOSDSMKROESPNNANDQPEDLIQGCOSQSDSVLRMSDPHTTNKLYPKDRLVDVSI	1033
Db	1388	SVSSLDSEFSRSLASSVOSEP-----CSGAWSGIISPDLDLPDSQGMWPMPSR	1434
Qy	1034	RSPETKPKPKNSIENDIEEKVDSLARGQEPYKSKTKGKNDPIIDSKLASAP-----	1085
Db	1435	-SKTPPPPOTAOTKRE-----VKNKAPPAEKRESG-----PKOAAVAAVOROVY	1480
Qy	1086	-DKGTILI-----CKNK-----KPLIOESDITV-----SSSKSALHSSV	1118
Db	1481	LPADNLDILHFATESTPDGFCSSSLSLALSDLEPFIQDVLEIMPPYOENDNGNTESEQ	1540
Qy	1119	PKSTINDOAPPLSRNADPBEKLCQDSENSSTLENSDVIYSDSSSEDMITVONSNESTIE	1178
Db	1541	PKSENEQO-----EKEAEKTIIDSEKD-----LDDSDDD-----	1569
Qy	1179	QFRRREDOVELEBLCKEIYSGESTGNCEDRLVKFTENGANKRPSQOKKLEERPYNKCS	1238
Db	1570	-----DIELE-EC-IISA-----MPTKSR-KAKKPAQOTASKLPPPAKRP	1609
Qy	1239	QI-----KIKNTDKKNNEERSEKKGORTSTFOJNGDKNPKITLYGECLEKIESESRYVG	1295
Db	1610	OLPURYKLLPSQNRLOPQKHVSFTPG-----DDMPRYV-----CV-----EGTPINF	1650
Qy	1296	NVEPKVNNIKIIPENDIKSLYKESAIKRPFIJNGVIMDMFERKSSFTKSHLSSDAE	1355
Db	1651	STATSLDILTTESPENL-----AAGBGVGAQOSEFEKRODTIPEGR-STDAQ	1700
Qy	1356	GNYRDSLETLPTSKES-----DSTQTTTPSASC	1364
Db	1701	GKGTSSV-TTPELDONKAEEGDILACINSAMPKGSHPFVKKIMQVQAASASSAP	1759
Qy	1385	ESNSVNOVEDMEIETSEVKVYTS-----SPTSEESNSLN	1420
Db	1760	NKNOLDCK-KAFTSPVKPIQONTEYRFRVRKNADSKNNLNAERYFSNKKDSKONLN	1817
Qy	1421	DFIDENGLPIKNNENYGE-----SKRK	1443
Db	1818	NSKOFNOKLRNNEDRVGSAFDPSPHHYPIBETPYCFGRNDSLSLDPEDDDDVLSSEK	1877
Qy	1444	TVTTEVTMTSTVATESKIYIKKEKDKOTVVSSTENCAKSTVY-----TTTTV	1493
Db	1878	---AELKKAENKESSEKXVSHTELSNOSASAKTOAIKOPINGORPILOKOSTPQ	1934
Qy	1494	TKLSTPSTGSDVILSVKEQSKVVTVTYTDLSITTTGGVLSMNVSKSYSTRDKVLKX	1533
Db	1935	SKSRIIDPRGAATD---EKLONFALIENTPVCFS---HNSLSLSLDIOENNKENEPIKE	1968
Qy	1554	FSRP-----KTRSGTALPYSRK-----FVTKSTKKSIFVLEPNDUL-----	1589
Db	1989	TEPDGSGEESKPPASQYAKRSNHVDTPVCFERNSSLSLSDSEDDLDLOCISSAPK	2048

QY	1590	-KTLARKGCIREFVFNYNAP-----ALDIWFPSPRPFFIGT-----WRX	1630
Db	2049	KKKSSRLTKGDNE-----KHSPPRMGGIIGEDLTLDLKIDRDPSEHSGISPENFDKA	21020
QY	1631	RLOGVNKLACGLSLMLRLMLSLMDMAKNVPDGGSTFETSEFEITTEILIRKRDVCP	16900
Db	2103	IQEGANSTIV-----SSLHQAAAACL-----SHQASSDSO---STLSLKS---	21399
QY	1691	YGIRFEYCIRKLIICPIGVPEPTEPTPOKGRSSALPKRPETPKOTGPVLIETWAE	17506
Db	2140	-GI-----SLGSFHLLTDQOEKEPPTSNGKR--ILKPGEKST-----LETKKIE	21810
QY	1751	EE-----LELMERAFERBEVEKEKAQVEDQAARLRDOQR-----TVIAT	17911
Db	2182	SESCKIQKGKKVVKSLITTGKVRNSN-----ISGOMKOPLDAMNPISRGRTHIH	22320
QY	1792	STTSPTSTSTTIPPAOKWVAAPSQSVTTGTAMVLITTKVSPARYVTPOQNNPHQTAT	18511
Db	2233	PCVRNSSSTSTSPSKKPPLKTAKSPSGQYATTTPSGKAPSV-----	22777
QY	1852	VWKQOSMSGVVQQAVQLVLIIPSNTSQSQTFTSPQRTAATVTRPNSTSGSGTYSNQ	19111
Db	2278	-----KSLSPVANQTOIG-----GSKR-----APRSKGRDSTPS--	23099
QY	1912	VITGPQINRGMTVIRLPLOOSTLGKALIRTPVMVOPC-----APOQVMQIIINGOPSTAV	19670
Db	2310	-----RPAQQPLSRPIIOBGRNS-----ISPRNGIISPPMKLSQLPTSTSPSTA-	23533
QY	1968	SAPTNTVSPTGOAKSLTS-----ATSTSNIQSSASOPRRPQOGQVKLYMAQLTQLTQ	20180
Db	2354	--STKSSGSGKMSYSTPGRMOSQOULIKOTGLSKMASSIPRESAS-----KGINMOINN	24050
QY	2019	GHGCGOGITVVYIQOGGOTTOLQ-----LIPOGYTVLBGGQODLMOAMPNGTVQFL	20711
Db	2406	GNAGANKVEELSRMSTSSGSEDSDREBPVLVNOSTFIKEAPSPILRRKLESAS----	24600
QY	2072	FTPPLATTATTTSTTTTSTTAAGTGEOROSKLSP-----OHQVHD-----KTLPPAQ	21200
Db	2461	FESLSPSSRAPSPTRSOAQRPV-----LSPILPMXLSTHSSVQAGGMRKLPNL	25110
QY	2121	SSSYGPAKAQPOTAQSPARPOTOPQSPAQPEVQTOPEVQTOYTIVSSHVSEAOPTHAQ	21800
Db	2511	SPITLEVNDGRAPKRDHDIARSHSESPEURLPINRSGTWKRE-----HSKHSLSLPLVSWRRTG	25670
QY	2181	SSKPVAAQSPQSNVOGOSPVPNVQSPSORIRNRSPSOLS-PGQOSQVQTTISOPIRIQ	22399
Db	2568	SSSSSTLSKASESESKAKSEDEKHVNSTLSG--KQSKENOVSAKGWTKIKENEESP-----	26211
QY	2240	PHTSLQIPSGOPQSPQPOVQOSTQTLSSQQT	2270
Db	2622	-----TNSTQTYVSSGAT	2634
 RESULT 14 US-08-449-731-7 Sequence 71, Application US/08449731 Patent No. 6413727 GENERAL INFORMATION: APPLICANT: ALBERTSEN, HANS ANAND, RAKESH CARLSON, MARY GRODEN, JOANNA HEGE, PHILIP J. JOSLYN, GEOFF KINKLER, KENNETH MARKHAM, ALEXANDER F. NAKAMURA, YOSUKE THLIVERIS, ANDREW TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC GENE IN COLORECTAL CANCER IN HUMANS NUMBER OF SEQUENCES: 102 CORRESPONDENCE ADDRESS: ADDRESSEE: Banner & Allegretti, LTD			

STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7
Query Match 2.4%; Score 337; DB 4; Length 2842;
Best local Similarity 17.1%; Pred. No. 2.5e-11;
Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;
QY 387 EIHRRHDITFDLTKKRGSKSLAA-ANDEILLESIRAKKGDIDNVKSPETKDKN--- 442
DB 801 DTKRHDDNRSDNFT---GMTVLSPYLNTVTLPLSSSSSKGSLDSSRS---EKDRLER 853
QY 443 -----ETENDSKDAEKNREFEEDOS-----LEKDDDKTPDDDEPOCK----- 480
DB 854 ERGIGLGNHPATENPGTSSKRLQISTTAQAIKAMEEVSALHTSQEDRSSGTTLELC 913
QY 481 -----SEVGF-KSEKSN-----GELSESPGAG 502
DB 914 VIDERNAIRSSAAHTHSNTYNTFCSENSNRICSMFYAKLEYKRSSNDLSNVSSSDG 973
QY 503 KGAGSSTRITTRIRNDPKLSQAKS---QVAAAHAHANKL-FREGKEVLVYN----- 551
DB 974 KRGQKPSLESSEDESKFCSYGYRADLAHRKIHSAHMDMDGDLDTPINSLKYSDE 1033
QY 552 ---SOGELISRLSTKEVIMKGINNNYFKLGOEGKRYRYHNOYSTNSPALNKHQIHREDHK- 608
DB 1034 QUNSGQSSQSONRMARPHIILEDIKOSEQORSR---NOSTT---YPVYTESIDDKHLNF 1088
QY 609 RRRHLAKFCULTPAGEFKMNGSVHGKSVLTITSLRLITOLENNIPSSFLHPNVAHRANW 668
DB 1089 OPHFGQOCECVSPYRSRGANGS-----ETNRVGS-----NHGINQ 1122
QY 669 IKAVQCSRPREFALALILECAVAVVMLPIVREFLGHTRLHRMTSIREKEKAKKE 728
DB 1123 NVSOSLCQED-----DYEDDKPTNVSE 1144
QY 729 KROEEETMQATWVKYTFPVKHWKQKGEYRVYGGGWSMISKTHYRVFVKPLPGNT 788
DB 1145 KXSEEL-----QHEEERPTNY----- 1161

QY 789 NNAYRKSLEGTKNNDENNMES-----DKRCSRPPKIKITIEPSEKD 831
DB 1162 SIKYNEE-----KHHVQPIDISLAKYATDIPSSQKQSFPSKSSGSSKTEHMSSENT 1217
QY 832 EVKSDAKAGADONEMDISKITEKKDQVKEILSDS-----DKP-CKEEPM 877
DB 1218 STEPSNAKR---ONOLHPSASGRSGOPKAKATCKVYSINOETIQYCVEDPFCPSRCS 1274
QY 878 EYVDDMKTESHVACQSSQ-----VDVNVSEGHRLT-----SYKKTKKS 919
DB 1275 SLSSLSAEDICNQTTOEADSANLTQIAEIKETIGTNSAEDPVSEVPAVSQHPRTKS 1334
QY 920 KLDGLERRIKOPTLEKORLEKIKLEGIKIGKSTNNSKN-----LSESPVI-TRAK 973
DB 1335 RLOG-----SSLSPSNAHKAVEFSSGAKSPSKGAQTPKRPPEHYQETPLMNSRGT 1387
QY 974 EGCQSDSMROBOSPMANNOPEDLIGCCSOSDSSVLRMSDPSTTKKLYPKDRVLDVSI 1033
DB 1388 SVSLSPFESRSIASSVOSEP-----CSQWVSGIISPDLPSDLPSPQTPMPSPR----- 1434
QY 1034 RSPETCPKONSLENDIEEKVSDLASRGQEPYKTKGNDFFIDSKLASAD----- 1085
DB 1435 ---SKTPPPPTQATKRE-----VPKNKAPYAEKRESG-----PKQAAVNAVORVOY 1480
QY 1086 ---DIGTLI-----CKNK-----KPLIOESDPIV-----SSSKALHSSV 1118
DB 1481 LPDADTLIHPATSTPDGFCSSSLALSIDEPIQKVDLRLMPVQNDNENEFESQ 1540
QY 1119 PKSTNDRADTPLSRAMDEBGLACDSESNSTLENSDDTVSIQDSEEDMIVQNSNEISE 1178
DB 1541 PKESNENO-----EKEAEKTIIDSEKD-----LDDSDDD----- 1569
QY 1179 QFRTGQDVEVLEPLCELVSGESTNCEDRLPVKGTSEANGKKPSQOKKLEBPVKNCS 1238
DB 1570 -----DIEILE-EC-IISA-----MPTKSR-KAKKPAQTAKLPPVAKRS 1609
QY 1239 QI---KIKNTTKKNNESESEKKGRTSTFOINGDKNKPKIYLKGECLKEISESRVVG 1295
DB 1610 QLPVYKLPSONMLQOKHVSFPFG-----DDMPRYV-----CV-----EGPINF 1650
QY 1296 NVKPKNNINKIIPENDISLYKESAIRPPIINGDIYMEDEFERNSEKSHLSSDAE 1355
DB 1651 STATSLDLTIESPBNEL-----AGEGVGAQSGEFERKDTIPIEGR--STDEAQ 1700
QY 1356 GNYRDSLETLPPSTKES-----DSTQTTTPSASCP 1384
DB 1701 GGTTSV-TIPELDNKAEBGLIACINSAMPKSKSHKPFYKKTIMQVQOASASSAP 1759
QY 1385 ESNVNOVEDMEIETSEVKKVTs-----SPITSEESNLIN 1420
DB 1760 NKNQOLDGKR--KKTSPVAPIPONTETRTVRKNADSKANLNAERYFSNKKSKONLKN 1817
QY 1421 DPLDENGILPINKNENNGE-----SKKK 1443
DB 1818 NSKDFNDKLPNNEDRVKGFAPDSPHHYPIEGTPYCFGRNDSLSPEDDDVDLSREK 1877
QY 1444 TYTTEVYTTMTSVATSEKSVIVKEKGDQOTVVSPTNCAKSVYT-----TTTTTV 1493
DB 1878 ---AEIRAKKKESEAKYTSHTELTSNOQSANIKOIAIKOPIINGOPRIILOKOSTPQ 1934
QY 1494 TKLSTPSTGSDVILSVKQSKTYVTTVYDLSITTTGGTLVTSMTVSKYSTRDVVKLMK 1553
DB 1935 SSKDIPRGAAND---EKLOFAIENTPVCFS---HNSLSLSLSDIDQENNNKKEPIKE 1988
QY 1554 FSRP-----KTTSGTALPSYRK-----FVYTSKKSITVLPNDL----- 1589
DB 1989 TEPPOSQEPSPKQASGYAPKSFHVEDTPVCFSRNSSLSSLSDIDEDDLOECISSAMP 2048
QY 1590 -KILARKGIREVPYFNVAKP-----ALDIMYPSRPRTFGIT-----WRY 1630
DB 2049 KKKPSRLKGDNE-----KHSFRNMGILGEDTLTDLKDIOBPSEHGSLPSENFPDKA 2102


```

QY 926 ERIKOTLEKORLEKIKLEGKIGKSTNSSKN-----LSESPVI--TKAKEGQSD 979
Db 1340 -----SSLSSASAKHKAVERPPSGAKSPSKGCAOTPKSPPEHYOETPLMSRCSVSSLD 1394
QY 980 SMROECPNANNDQPEDLIOGCSOSDSSVLRMSDPSTHTNKLYPKDVLADVSIRESPTK 1039
Db 1395 SFEERSJLASSVQSEPR-----CSCMGSIISPSDLPSDPOQTMPSR-----SKTP 1439
QY 1040 CPKONSTIENDIEEKVSLASRGQEPYKSKTKGNDFFIDDSKLASAD-----DIGT 1089
Db 1440 PPPOTAOOTKRE-----VPKNKAPTAEKREBSG-----PKQAAVNAVORVOYLPPADT 1487
QY 1090 LI-----CKNK-----KPLIOESDTIV-----SSSKALSHSPKSTND 1124
Db 1488 LLIHATSTPGEFSCSSLSALSLDEPFIQKDVLRIMPYQVENDNNGENESBPKRSNE 1547
QY 1125 RDATPLSRAMDEBKLCODESSENSTLENSDPTVSIODSSEDMIVQNSNESISEQFTRE 1184
Db 1548 NO-----EKEAEKTIIDSEKD-----LLDSDDD----- 1570
QY 1185 ODVEVLEPLKCELVSGESTGNCERLRYKGTENGKKPQOQKLEERPVKCSNOI---K 1241
Db 1571 -DIFLE--EC-TISA-----MPTKSSR--KGKPPAOTASKLPPVPVARKPSQLEPVK 1616
QY 1242 LKNTTDKKNENRESEKKGORTSTFOJNGKDNKPKIYLGCELEISESRVSGNVBPKV 1301
Db 1617 LLPSONRLQPOKHVSFTMG-----DDMPRVY---CV-----EGPFINSTFATSL 1657
QY 1302 NNINKIIPENDIKSLTYKESAIRPPINGDVIMEDPNERNSSETSHLSSSDAGNTRDS 1361
Db 1658 SDLTIESPPNEL-----AAGEVNRGAQSGEGERKDTIPTGR--STDEAGGGTSS 1707
QY 1362 LETLPSTKES-----DSQOTTPSACESSVNV 1390
Db 1708 V-TIPELDONKAEBGDILAECSNAMPKKGKSHKPPRYAKKIMDQYOQASASSAPKNOLD 1766
QY 1391 QVEDMEIETSEVKKYVS-----SPITSEESNLSDFDEN 1426
Db 1767 GKK--KKPTSPVKPIPOUTEYRTVRKKNADSKNNLNAERVSFSDNKKDKQKLNNSDEN 1824
QY 1427 GLPIKNENVGE-----SKRKTIVITEV 1449
Db 1825 DKLPPNEDRVGSEFAFDSPHHYTPLEGTPYCFSRNDSLSDFDDDVDLSREK---AEL 1881
QY 1450 TTMSTVATESKTIVKEGDKOFVVSSTENCASVT-----TTTTVTKLSTP 1499
Db 1882 RKAKEKSEAKVYSHTELTSNOSANKTOALAKOPINRGPKPILOKQSTFFPOSSKDIP 1941
QY 1500 STGGSVDIISVKEOSKTVVTTVTDSLTGGLVTSMTVSKKEYSTRDKVKLMKFSRP-- 1557
Db 1942 DRGAATD--EKLONFALIENTPVCS--HNSSLSLSLDIDQENNNKENEPIKETEPDS 1995
QY 1558 -----KTKRSGTALPSYK-----FVTKSKTSIFVLJPNDDL-----KKLAR 1594
Db 1996 OGEPSKPOASGAPKSFVEHEDTPVCFSRNSSLSLSDSEDDLLOECISSAMPKKKPSR 2055
QY 1595 KGIREVYFNYNAPK-----ALDIWYPSPRPRTGIT-----WRYRLQTVK 1636
Db 2056 LKGDME-----KHSPRMMGGILGEDLTLDKIDRPSHGLSPDSSENFDMWKALQEGAN 2109
QY 1637 SLAGVSLMLRLIMASLRMDDMAKVPFGGSGSTRFETSETETTTETITKRDOVGYPGIRFE 1696
Db 2110 STV-----SSIHQAAAAACL-----SROASSDS-----SILSLKS---GI--- 2142
QY 1697 YCIRKIICPIGVETPKETPTPQKGLSSALRPKRPEPKOTGPVILLETWAAEEE--- 1752
Db 2143 -----SLGSPFHLTPQOEKPTSNKGR--LIKPEKST-----LETKKISESKGI 2188
QY 1753 -----LELMEIRAFEREREKEAQAVEQOAKKRLQOKP-----TVIATSTTSP 1797
Db 2189 KGGKVVYKSLITGKVRNS-----ISQOMKQLOANMPSLSRGRTHIHIPGVANS 2239
QY 1798 SSTSTISPAOKVWVAPISGSVTTGTKNVLTTKVGSFATVTFQONKNFHOTFAIWKOG 1857

```

```

Db 2240 SSTSPVSKKGPPLKTPPAKSPSEGCOTATTSPRGAKPSV-----K 2279
QY 1858 SNNGVVOVQOKVIGITIPSSGTGSOQFTSFQORTAVITIRPMTSGSGGTTNSQVITGP 1917
Db 2280 SELSPARQOTSOQ-----GSSK-----APSRSGRSTPS----- 2310
QY 1918 IRPGMTVIRTPLOQSTLGAIRITPVAVOFG-----APQVMTQIIRGQFVSTAVSAPNTV 1973
Db 2311 -RAQOPLSPRIOSPGRNS-----ISGRNGISPPMKLSQLPRTSSPSTA---STK 2357
QY 1974 SSTPGKSLTS-----ANSTNSIOSSASQPPRPQOGQVKLYMAQULTOLTQHGHNQ 2024
Db 2358 SSGSGKMSYSTSPGRQMSQONLTKOTGLSKNASSIPRSESAS-----KGLNOMNNGANK 2412
QY 2025 GLTVIVUGQOQTIGOLQ-----LIPQGVTVLPGGQQLMQAAMPNGTVQFPLFTPLAT 2077
Db 2413 KVELSRMSSTKSSGSDSNDSERPVLVROSTFIEKAFSPTLRKLRESAS-----FESLSP 2467
QY 2078 TATTAATTTTTVSTTAAGTGEOROSKLSF-----QMQVHOD-----KTLPPAQSSSVGP 2126
Db 2468 SSRPASPTRSQAOPTPV-----LSPSLPMSLSSTHSSVQAGGWKRLPMLSPTIEX 2517
QY 2127 AKAOPTOAPSARPQOTOPQSPAPQPEVOTQPEVOTQTVSSHVPSBQPTNAOSKPOV 2186
Db 2518 NDRGPARKRHDIAKSHSESPELRPLINSGTWKRE--HSHKSSLPVSVTWRTGSSSIL 2574
QY 2187 AOSQOPQSNVQSPVAVQSPQTRIRPSPQSL--PGQOSQVQVTTQSOPPIPIQPTSLQ 2245
Db 2575 SASSESEKKSDEKHHVNSISGT--KQSKENGVSAKGTMRKIKENEFSP----- 2622
QY 2246 IPSQOGPOQOPQVQVSTQTLSSGQT 2270
Db 2623 -----TNSTSOVSSGAT 2635

```

Search completed: November 20, 2002, 16:34:16
 Job time : 142.024 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:01 ; Search time 99.1487 Seconds

(without alignments)
6041.216 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971
Sequence: 1 MVSEEEEDGDGAETQDSE.....KIKGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14170	94.6	2781	4 09UG2	09ug2 homo sapien
2	2771.5	18.5	2669	5 09SVB8	09svb8 drosophila
3	2705.5	18.1	645	11 08VDN7	08vbn7 mus muscula
4	1488.5	9.9	976	5 09W0T0	09w0t0 drosophila
5	1368.5	9.1	412	4 09H5E0	09h5e0 homo sapien
6	854.5	5.7	1711	5 04S4O9	04s4o9 caenorhabdi
7	848.5	5.7	1713	5 09S4U8	09s4u8 caenorhabdi
8	692.5	4.6	1022	5 0960Y3	0960y3 drosophila
9	661.5	4.4	997	5 09W0T1	09w0t1 drosophila
10	621.5	4.2	2768	5 09VC00	09vc00 drosophila
11	613.5	4.1	5476	5 09AJ17	09aj17 drosophila
12	613.5	4.1	5533	5 09VPL2	09vpl2 drosophila
13	613.5	4.1	5554	5 09NHN1	09nhn1 drosophila
14	613.5	4.1	5560	5 09VPL1	09vpl1 drosophila
15	608.5	4.1	5533	5 09UC63	09uc63 drosophila
16	593	4.0	17352	5 095YM2	095ym2 procamburus

17	571.5	3.8	16215	5 09NFS3	09nfs3 drosophila
18	567	3.8	6815	5 0917U4	0917u4 drosophila
19	562.5	3.8	3257	5 09V736	09v736 drosophila
20	561	3.7	3080	5 09VRY3	09vry3 drosophila
21	551.5	3.7	2897	5 09VTD9	09vtd9 drosophila
22	546.5	3.7	5327	5 076891	076891 drosophila
23	518.5	3.5	2081	10 09LH98	09lh98 arabidopsis
24	513.5	3.4	2465	5 022463	022463 caenorhabdi
25	508	3.4	2910	10 09FND5	09fnd5 arabidopsis
26	503.5	3.4	4498	5 09W223	09w223 drosophila
27	496	3.3	6994	5 017343	017343 caenorhabdi
28	490	3.3	5170	5 017490	017490 caenorhabdi
29	489	3.3	2402	2 09AER7	09aer7 staphylococ
30	474	3.2	3766	5 09VFL1	09vfl1 drosophila
31	463.5	3.1	3507	5 023587	023587 caenorhabdi
32	462	3.1	2586	5 09VTK8	09vtx8 drosophila
33	460.5	3.1	3484	5 P91257	P91257 caenorhabdi
34	456.5	3.0	3261	4 09Y556	09y556 homo sapien
35	456.5	3.0	3664	4 096T58	096t58 homo sapien
36	453	3.0	3551	11 09JPS4	09jps4 mus musculu
37	451	3.0	5085	11 09JKS6	09jks6 rattus norv
38	449.5	3.0	3111	5 09VH10	09vh10 drosophila
39	449	3.0	2283	2 08V099	08v99 staphylococ
40	444	3.0	2977	5 09VAP9	09vap9 drosophila
41	441.5	2.9	2083	5 09NA35	09na35 caenorhabdi
42	441.5	2.9	5120	13 09PU36	09pu36 gallus gall
43	440.5	2.9	3576	11 09Q2W2	09q2w2 mus musculu
44	440	2.9	2271	16 0980Y4	0980y4 staphylococ
45	440	2.9	2471	13 09DDN8	09ddn8 xenopus lae

ALIGNMENTS

RESULT 1					
09UG2		PRELIMINARY;	PRT;	2781 AA.	
AC 09UG2;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)					
DE Bromodomain PHD finger transcription factor.					
GN BPTF.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE-20130111; PubMed-10662542;					
RA Jones M.H., Hamana N., Shlman M.;					
RT *Identification and characterization BPTF, a novel bromodomain					
transcription factor.;					
RL Genomics 63:35-39(2000).					
DR EMBL; AB032251; BAA89208.1; -					
DR HSSP; O92831; 1B91.					
DR InterPro: IPR001487; Bromodomain.					
DR InterPro: IPR004022; DDT_dom.					
DR InterPro: IPR000561; EGF-like.					
DR InterPro: IPR001965; ZnF_PHD.					
DR Pfam: PF00439; bromodomain; 1.					
DR Pfam: PF002791; DDT; 1.					
DR Pfam: PF00628; PHD; 2.					
DR PRINTS; PRO0503; BROMODOMAIN.					
DR SMART; SM00297; BROMO; 1.					
DR SMART; SM00249; PHD; 2.					
DR PROSITE; PS00633; BROMODOMAIN_1; 1.					
DR PROSITE; PS00014; BROMODOMAIN_2; 1.					
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.					
SO SEQUENCE 2781 AA; 311210 MW; 783BE0C1C4C3CFBE CRC64;					
Query Match	94.6%;	Score 14170;	DB 4;	Length 2781;	
Best local Similarity	95.1%;	Pred. No. 0;			

Matches 2766: Conservative 2; Mismatches 13; Indels 126; Gaps 1;	
QY 1 MYSEEEEDDADAEETDSEDDEDEEDDDSDYPEEMDDDDDAASYCESSFRSHST 60	
Db 1 MYSEEEEDDADAEETDSEDDEDEEDDDSDYPEEMDDDDDAASYCESSFRSHST 60	
QY 61 YSSTPGRRKPRVHRPRSPILIEEKDIPLEFPKSSSEDLMPNEHIMNYAIEVLRNFTV 120	
Db 61 YSSTPGRRKPRVHRPRSPILIEEKDIPLEFPKSSSEDLMPNEHIMNYAIEVLRNFTV 120	
QY 121 LRLSPFRREDCAALVSOEECTLMAEMHYLLKAVLRBEDTSNTTGPADLKDSVNSTLY 180	
Db 121 LRLSPFRREDCAALVSOEECTLMAEMHYLLKAVLRBEDTSNTTGPADLKDSVNSTLY 180	
QY 181 FIDGATPEVLRVYCESKEKEHYHVLPLYOEAEDYRGPVENKIKYLOFVLDTNTIARE 240	
Db 181 FIDGATPEVLRVYCESKEKEHYHVLPLYOEAEDYRGPVENKIKYLOFVLDTNTIARE 240	
QY 241 ELMSEGVIOYDHCRCVKRLGDLCCETCSAVYHLECVKPPLEEVPEDEMOCCEVCAHKV 300	
Db 241 ELMSEGVIOYDHCRCVKRLGDLCCETCSAVYHLECVKPPLEEVPEDEMOCCEVCAHKV 300	
QY 301 PGVYDCAVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIEEDTENEKEKIWTYSTVQ 360	
Db 301 PGVYDCAVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIEEDTENEKEKIWTYSTVQ 360	
QY 361 LAELIDCDKDYWEAELOCKILEEMKEEIHHRMDITEDLTNRKARSKNSFLAANAEILLES 420	
Db 361 LAELIDCDKDYWEAELOCKILEEMKEEIHHRMDITEDLTNRKARSKNSFLAANAEILLES 420	
QY 421 IRAKKGIDIDNYSPEETEKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDDEPEQK 480	
Db 421 IRAKKGIDIDNYSPEETEKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDDEPEQK 480	
QY 481 SEEPTEVODKNSVANSNGDVTNTASETSPSEGRSPVGLSETPDSSNAEKKVASL 540	
Db 481 SE----- 482	
QY 541 PQDVPPEENKTCESNTSATTTTSIQPLNLSNSSSELNSSQSESAKADDPENGERESHT 600	
Db 483 ----- 482	
QY 601 PVSIOEELVGDFTSEKSTGELSESFGAKGASGSTRITRLRNPDSKLSQKSOQVAAA 660	
Db 483 -----VGDFFSEKSENGELSESFGAKGASGSTRITRLRNPDSKLSQKSOQVAAA 534	
QY 661 HEANKLFKEGKELVYVNSOGELSRSTKEEVIYMKGINNYEKLGOEGRYRYHNOYSTNS 720	
Db 535 HEANKLFKEGKELVYVNSOGELSRSTKEEVIYMKGINNYEKLGOEGRYRYHNOYSTNS 594	
QY 721 FALANKHOREDDHDKRRHLAHLFCULTPAGEFKWNSVHSGSVLJTSTLRLLTTOLENNIPS 780	
Db 595 FALANKHOREDDHDKRRHLAHLFCULTPAGEFKWNSVHSGSVLJTSTLRLLTTOLENNIPS 654	
QY 781 SFPHHNAASHRANMTKAYONGSKPREPALALALECAVKPVVMPPIWBEFLGHTRLHMT 840	
Db 655 SFPHHNAASHRANMTKAYONGSKPREPALALALECAVKPVVMPPIWBEFLGHTRLHMT 714	
QY 841 STEREEKKVKKKKEKOESEETMOATWVKKYTFPVKHOMVOKOGEYVNTGCGSWIASK 900	
Db 715 STEREEKKVKKKKEKOESEETMOATWVKKYTFPVKHOMVOKOGEYVNTGCGSWIASK 774	
QY 901 THVRFVFKLPGNTVNVNRKSLGKTKNNMDENMDESARKCSRSFKIKIEPDSKDEVK 960	
Db 775 THVRFVFKLPGNTVNVNRKSLGKTKNNMDENMDESARKCSRSFKIKIEPDSKDEVK 834	
QY 961 GSDAAKADONEMDISKTTEKKODDYKELLDSDSKPCKEEMVEYDDMKTESHYNCOES 1020	
Db 835 GSDAAKADONEMDISKTTEKKODDYKELLDSDSKPCKEEMVEYDDMKTESHYNCOES 894	
QY 1021 SOVDVNVNVEGFHLRTSVKKTTSKSLDGLERRIKOFTLEKOKLEKIKLEGGIKGIGK 1080	
Db 895 SOVDVNVNVEGFHLRTSVKKTTSKSLDGLERRIKOFTLEKOKLEKIKLEGGIKGIGK 954	

QY 1081 TSTNSKLNSESPLYTKAKEGCOSDSMRQOSPANNDDPEDLIOGCSQSDSSVLRKMSDP 1140	
Db 955 TSTNSKLNSESPLYTKAKEGCOSDSMRQOSPANNDDPEDLIOGCSQSDSSVLRKMSDP 1014	
QY 1141 SHTTNKLYPKRDVLDVYSINSPETKCPKONSIEDEEYKVSOLASRQOETKSKTKGNDP 1200	
Db 1015 SHTTNKLYPKRDVLDVYSINSPETKCPKONSIEDEEYKVSOLASRQOETKSKTKGNDP 1074	
QY 1201 FIDDSKLASADIGTLCKKKKPLIOESDPTIYSSKSALHSSVPKSTNDRDPTPLSRAM 1260	
Db 1075 FIDDSKLASADIGTLCKKKKPLIOESDPTIYSSKSALHSSVPKSTNDRDPTPLSRAM 1134	
QY 1261 DFEKGLCDSESNSTLNSSDTVSIOSSSEDMIVONSNEISBOFTREODVEVLEPLK 1320	
Db 1135 DFEKGLCDSESNSTLNSSDTVSIOSSSEDMIVONSNEISBOFTREODVEVLEPLK 1194	
QY 1321 CELVSGSTGNCEDRLPVKGTENGKRPSSQKLEERPVKCSQIKLKMTTKKNENR 1380	
Db 1195 CELVSGSTGNCEDRLPVKGTENGKRPSSQKLEERPVKCSQIKLKMTTKKNENR 1254	
QY 1381 ESEKKGQRTSFOINGKDNKPKIYLKGECLKEISESRVSGNVEPKVNNINKIIPENDIK 1440	
Db 1255 ESEKKGQRTSFOINGKDNKPKIYLKGECLKEISESRVSGNVEPKVNNINKIIPENDIK 1314	
QY 1441 SLTVKESAIRPFINGDVIEMEDFENRNSSETKSHLSSSDAEGNROSLETLPTKESDST 1500	
Db 1315 SLTVKESAIRPFINGDVIEMEDFENRNSSETKSHLSSSDAEGNROSLETLPTKESDST 1374	
QY 1501 OTTPSPASCPESNVNOVEMELETSEVKYTSPTISEESNLSNDFIDENGPIPKNE 1560	
Db 1375 OTTPSPASCPESNVNOVEMELETSEVKYTSPTISEESNLSNDFIDENGPIPKNE 1434	
QY 1561 NVNGESKRKYTYTEVTMTSTVATESKTVYKVEGDKQTVVSSSTENCAKSTVTTTTVT 1620	
Db 1435 NVNGESKRKYTYTEVTMTSTVATESKTVYKVEGDKQTVVSSSTENCAKSTVTTTTVT 1494	
QY 1621 KLSPTSGGSDVILISVKEOSKTVVTTVTDSLTGTLVTSMYSEKYSTRODKVLMKF 1680	
Db 1495 KLSPTSGGSDVILISVKEOSKTVVTTVTDSLTGTLVTSMYSEKYSTRODKVLMKF 1554	
QY 1681 SRPKTRSGTALPSYRKFTVSTKKSIFVLPNDLKLARKGIREVPYNYNAKPLADI 1740	
Db 1555 SRPKTRSGTALPSYRKFTVSTKKSIFVLPNDLKLARKGIREVPYNYNAKPLADI 1614	
QY 1741 WPPSPRPTGRTWRKRYRLQTVKSLAGVSLMDRLMASLRMDMAKVPFGGSGTRTETSE 1800	
Db 1615 WPPSPRPTGRTWRKRYRLQTVKSLAGVSLMDRLMASLRMDMAKVPFGGSGTRTETSE 1674	
QY 1801 TEITTELIRKRDVGPYIGIREYCIRKILCPIGVPEPKETPTPQRKGLASSALRPKRE 1860	
Db 1675 TEITTELIRKRDVGPYIGIREYCIRKILCPIGVPEPKETPTPQRKGLASSALRPKRE 1734	
QY 1861 TPKOTGPVLIETWVAEELMELETRAFAERERKKAQAVEQAKRLLEOKPVIATST 1920	
Db 1735 TPKOTGPVLIETWVAEELMELETRAFAERERKKAQAVEQAKRLLEOKPVIATST 1794	
QY 1921 SPTSSTSTISPAOKVNVAPISGSVTTGKMWLTJKVSPATVTFQONKNFHQTFAIWK 1980	
Db 1795 SPTSSTSTISPAOKVNVAPISGSVTTGKMWLTJKVSPATVTFQONKNFHQTFAIWK 1854	
QY 1981 QGOSNGVVOYQKVLGIIPBSTGTSOOTFTSFOPRATVYIRANTSGSGGTTNSQVIT 2040	
Db 1855 QGOSNGVVOYQKVLGIIPBSTGTSOOTFTSFOPRATVYIRANTSGSGGTTNSQVIT 1914	
QY 2041 GPOIRPGMTVIRTPLOQSTLGKALIRTPVNVOPAPQOVMQOIRRGOPVANSAPTVS 2100	
Db 1915 GPOIRPGMTVIRTPLOQSTLGKALIRTPVNVOPAPQOVMQOIRRGOPVANSAPTVS 1974	
QY 2101 STPGOKSLTSATSTNSIOSSASQPPRPOGQGVKLTMAQLTOLQGHGNGGLVAVIOGQ 2160	
Db 1975 STPGOKSLTSATSTNSIOSSASQPPRPOGQGVKLTMAQLTOLQGHGNGGLVAVIOGQ 2034	

QY	2161	QTTSQLOLIRQGVTVLPBPGQOLMOAMRNGVORFLPLATTTATTTASTTTTYYSTTAA	2220
Db	2035	QTTSQLOLIRQGVTVLPBPGQOLMOAMRNGVORFLPLATTTATTTASTTTTYYSTTAA	2094
QY	2221	GTGEGOROKSLSPOMOVYHODTLTPPAOSSSSVGAKKAPORAPASARPOQTOPSPADEV	2280
Db	2095	GTGEGOROKSLSPOMOVYHODTLTPPAOSSSSVGAKKAPORAPASARPOQTOPSPADEV	2154
QY	2281	QTOPEVOTQTTVSSHVPSEKOPTTHAASSKPOVAASQOPASNVOGSPVRVOSPSTRIRP	2340
Db	2155	QTOPEVOTQTTVSSHVPSEKOPTTHAASSKPOVAASQOPASNVOGSPVRVOSPSTRIRP	2214
QY	2341	STPQSLSPGQSOYQTTTSPRPIPTQHTSLQIPSGQOPQSPQVQVSSQTULSSGOTLNV	2400
Db	2215	STPQSLSPGQSOYQTTTSPRPIPTQHTSLQIPSGQOPQSPQVQVSSQTULSSGOTLNV	2274
QY	2401	SVSSPSPRQLOIOPOROVAVAPQLOOQVQVQVLSQISQOVVAIOAGQSGVPOQIKLOPI	2460
Db	2275	SVSSPSPRQLOIOPOROVAVAPQLOOQVQVQVLSQISQOVVAIOAGQSGVPOQIKLOPI	2334
QY	2461	QIOQSSAVQTHQIOMVTVVVOAASVOBQLORVQDLRQDQKKKKQOQIETINVTPSKLILY	2520
Db	2335	QIOQSSAVQTHQIOMVTVVVOAASVOBQLORVQDLRQDQKKKKQOQIETINVTPSKLILY	2394
QY	2521	EIIQOKVYVNMHNAVIELKKKSKMTAEENQRMIVCQVQVKTILDKIDKEKOAAKR	2580
Db	2395	EIIQOKVYVNMHNAVIELKKKSKMTAEENQRMIVCQVQVKTILDKIDKEKOAAKR	2454
QY	2581	KREESVEOKRSKONATLSLFLFKHEQOLRAELIKKRALLDDLOIEVOEELKRDQLTKK	2640
Db	2455	KREESVEOKRSKONATLSLFLFKHEQOLRAELIKKRALLDDLOIEVOEELKRDQLTKK	2514
QY	2641	EKDLMOAQATAVAAPCPPTVPLPAPAPAPPSPPPVQHTGTLSTPTLPVASQKKR	2700
Db	2515	EKDLMOAQATAVAAPCPPTVPLPAPAPAPPSPPPVQHTGTLSTPTLPVASQKKR	2574
QY	2701	EEEDOSSSKSKKKMMISTTSKEFKADTKLYCICKTPYDESKFYICQDRCQVMYHRCVGI	2760
Db	2575	EEEDOSSSKSKKKMMISTTSKEFKADTKLYCICKTPYDESKFYICQDRCQVMYHRCVGI	2634
QY	2761	LOSABELIDEVQPOCOSTEDATVTLPTLENDYBSLKVLSLOAHKMAPPLEPNDP	2820
Db	2635	LOSABELIDEVQPOCOSTEDATVTLPTLENDYBSLKVLSLOAHKMAPPLEPNDP	2694
QY	2821	DADPYGVIKEPMDLMTMEERVORARYEKLTEFVADMTKIFPNCRVYVNPSPFYQCAV	2880
Db	2695	DADPYGVIKEPMDLMTMEERVORARYEKLTEFVADMTKIFPNCRVYVNPSPFYQCAV	2754
QY	2881	LESFFVQKLKGFKASRSHNNKLOSTAS	2907
Db	2755	LESFFVQKLKGFKASRSHNNKLOSTAS	2781
RESULT 2			
ID	Q95VB8	PRELIMINARY:	PRT: 2669 AA.
AC	Q95VB8;		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Nucleosome remodeling factor large subunit NURF301.		
GN	El(BX) OR NURF301 OR CG7022 OR CG1135.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila;		
OX	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21468388; PubMed=11583616;		
RA	Xiao H., Sandalitzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,		
RT	Fu D., Wu C.;		
Dual functions of largest nurf subunit nurf301 in nucleosome sliding			

Query Match	18.5%	Score 2771.5	DB 5	Length 2669
Best Local Similarity	26.5%	Pred. No. 1,4e-102		
Matches 869	Conservative 384	Mismatches 920	Indels 1107	Gaps 99
Query	4	EEEEEDDGAETOSSEDEDEMEDEDDDDSYPREMEDDDDDASVCTSSSRSHSTYSS	63	
Db	99	EESEYHSGSFSGSESDKSDNEDMDLTPSDSDSEVANSSESFVSC---SFNQNV---	152	
QY	64	TPGRKRPVHRHRSPLLE-KDIPLEPKSSSEDLVMPREHIMNVAIYEVLRNFTVLR	122	
Db	153	---GRRPRRSEPPWLDGRROYALDLRPSDEDLFIATHTVLRALSTIEVLRRRRHRMR	209	
QY	123	LSPPRFEDFCAALVSQEOCTLMAEMHVVLLKAVLRBEDTSNTTFGPADLSDSVNSTLYET	182	
Db	210	LSPPRFEDFCAALVEGQALTLVEHIMLKALIREDAQGTNFGPLDQKDTVNISLYLI	269	
QY	163	DGMTPEVILRVYCESDKEHIVLPRQELADPRYGVEVKKIKIVLOLVQOFLTNLA	238	
Db	270	DSITWPEVILRVYCESDKEHIVLPRQELADPRYGVEVKKIKIVLOLVQOFLTNLSI	326	
QY	239	REELSEGVIQYDDHCRVCHKGLDLCCETGSAVYHLECVKPRPLEVPEDEMQCEVYAH	298	
Db	327	RVMYLOEGRINHDDICRCHRLGDLCCETGCAVYHLECVDPMDVPRMEDQGLGRSH	386	
QY	299	KVPGVTVCAELQKKNPYLRNHEPIGYDRSRKRYWFLNRRLLI EEDTENENKKIMYSTR	358	
Db	387	KVSGVVCVLPQEKQGVILRHSDLGVDHGRKRYWFIARIIFI-EDQENPT---CWYSTT	442	
QY	359	VQLAELIDCLDKDYNEAELCKLLEEMREIRIHHMDITEDLTAKAGSNKSFLLAANEFL	418	
Db	443	SKLLKLLSLRLAEELTRLHSQITERRBOIEOMKLTLLTLENHKTJKTSVI-----	494	
QY	419	ESIRAKGIDHWKSPLETERKDNFTENDSKAENKREEFQOSLEKDDTTPDDPEQ	478	
Db	455	-----ELDEAKNELLEKVLDD--EKGDGAKSEOSISTG	528	
QY	479	GKSEEPTEVGDKGNSVSNALGDNNTNATSEITSPSEGRSPVGLSETPDDSSMAEKVYAS	538	
Db	529	TKKOE-----	533	
QY	539	ELPQVYPREPKTCCSSNTSATTTISIQPLNLENSNSSLNQSOSAKARADPENGERS	598	
Db	534	-----	533	
QY	599	HTPVSIOEIVGDTFSEKSTGELSESFGAGSGASTRIITRLRNPDSKLSQLKSOQVAA	658	
Db	534	-----ECKWVR-----QKSNOLT	547	
QY	659	AAHEANKLFEGKEVLVYNSQELISRLSTKKEVIMKGNINNYFKLGOEKKYRVYHNOYST	718	
Db	548	-----NGLHFKLGMEGGFYNYNOYST	570	
QY	719	NSFALNKQHQRHDDKRRHLAKFCLLTPAGERKKWAGSVGKAVLITSLRLTLTIOLENNI	778	
Db	571	NPALANKQORNEBRKRRHLSHKFSLTIVASDFKMTGKITGMDNNITLTKQTLNFEESI	630	

Qy	779	PSSEFHPWMA\$BRANWIKAVOMCSKPREPFLALAILIECAVKKPVMLPIWREFLGHTRLHR	838
Db	631	AASFINTIMVWYKATIMWAAVNNARBPSEFVVLLEFQASLKSVEFANVWHEDLGHTTOR	690
Qy	839	M\$ISIEREKREKVKREKKO---EEETMQOATWYKTFPYKHQWKGGEERYATGCGW	895
Db	691	ITSAREERBKLEKREKREDEEERKLAFWIKTTLGLKHQWYMGKEELRVGONGW	750
Qy	896	SWIKST-----HYVFWPKLPGNTNVYRKSLEGTNMDENANDES\$DKRCSR\$KIKI	950
Db	751	LWLS\$SRRCGVARARAPLTHNRYVYHYTW---GEENDVNEI-----LVDRTOGR	799
Qy	951	EPDSKDEKVS\$DAKAGDQ--NMDSIKTEKKDQVYKELLDSD\$DKRCKEPEWVD	1008
Db	800	MOQSS\$NVDOQVCHYLPDQKKNKYLEDTE-----	831
Qy	1009	MKTESHVNOQESSOVADVNVSEGFHL--PT\$YKKT\$KSLGOLERRIKOFTLEEKRL	1066
Db	832	KIKGH-----IDVSKALNAGRIYYSKVARKSRDLDLDRKLALVEE-QMA	878
Qy	1067	EKIKLEGKIGIKGTSTN\$SKNLSSE\$FVITKAEQCSQSD\$MNOEQ\$PANNANDQEDLIQG	1126
Db	879	SKI-----	881
Qy	1127	CSQSD\$VLR\$SDPSHTNKLKYPRVDLDOVSIR\$P\$TCQPN\$STENDIEEK\$VDL\$R	1186
Db	882	---P\$M\$KPLV\$SONNTANS-----KOTLE-----KR	907
Qy	1187	GOEPT\$K\$TKGNDEFID\$SKLASADIGTLICKNKKPLIOESBPTV\$SS\$K\$ALH\$SVK	1246
Db	908	ILRLTEVQAK-----	918
Qy	1247	STNBDATPPL\$RANDEFGKLGCSQ\$ESNSTLENSQ\$DVSIO\$DSEEDMIVQ\$N\$ESIQF	1306
Db	919	-----GPANVLMELVNSLAKOIQVRLQ\$F-----	943
Qy	1307	RTREODVEVLEPLCELV\$GESTGNCEDRLP\$VGT\$EANGKRP\$OOKLEERP\$NKC\$DOI	1366
Db	944	---Q\$LNFAV\$FPC-----YTRKC-----NT\$N\$M\$V\$SQITQ-----MTC\$P\$PL	978
Qy	1367	KLKNTTDK\$N---NEKRE\$KKGQRT\$TFQ\$Q\$KNDK\$NPKIYKGECLKEI\$ES\$H\$V\$GN	1422
Db	979	CLQ\$ARAKKELL\$LRKAH\$T\$AGN\$SE\$T\$V\$A\$ILG\$AVK\$K\$S\$IL-----EOKLE\$GK	1028
Qy	1423	VEPK\$VNN\$INKIIPENDIK\$ILTVK\$ESAIRPE\$INGDVI\$MED\$FN\$RNS\$EMT\$K\$H\$LL\$SS\$D\$EG	1482
Db	1029	RE-----STQ\$V\$AVD\$D\$E\$E\$GK\$P\$AES\$AP\$LDL\$Q\$DEH	1059
Qy	1483	---NYR\$SL-----ETL\$P\$TK-----ESD\$TO\$TTP\$PSA\$C\$E\$S\$N\$VNO\$DEM-	1521
Db	1060	ARAHV\$P\$F\$D\$B\$ILTEC\$ILV\$DQ\$E\$CV\$T\$N\$KIKO\$E\$V\$N\$S\$G\$C\$NTT\$PDS\$N\$Q\$D\$SKIDY\$E\$S\$M	1119
Qy	1522	---ET\$E\$SVK\$V\$Y\$T\$SP\$IT\$SE\$E\$N\$L-----SND\$FID-----ENGLP	1555
Db	1120	VCSNVEI\$E\$T\$D\$S\$IVTG\$N\$G\$M\$A\$ED\$V\$M\$P\$G\$M\$R\$K\$K\$N\$O\$K\$K\$S\$Y\$IGT\$D\$V\$LD\$Q\$TL\$DK\$DIP	1179
Qy	1556	INKNENVNGE\$K\$R\$K\$T\$VTE\$V\$T\$T\$V\$T\$S\$T\$V\$AT\$E\$S\$K\$T\$V\$I\$VEK\$G\$Q\$T\$V\$S\$T\$E\$N\$C\$A\$S\$T\$Y\$T\$T	1615
Db	1180	L\$N\$K-----Q\$N\$R\$F\$IT-----ARV\$K\$R\$E\$C-----YK\$Y\$E\$T\$E\$N\$G\$N\$E\$R\$Y-	1214
Qy	1616	TTTVTKL\$T\$P\$T\$G\$S\$V---DIISV\$K\$Q\$K\$T\$V\$V\$T\$T\$V\$D\$SL\$T\$T\$T\$G\$T\$V\$T\$S\$M\$T\$V\$S\$K\$E\$Y\$T	1671
Db	1215	---Y\$T\$S\$P\$R\$G\$R\$V\$YL\$INDAK\$ILE\$Q\$A-----YKT	1240
Qy	1672	RDKV\$K\$LM\$K\$F\$R\$P\$K\$T\$S\$G\$T\$AL\$P\$Y\$K\$R\$F\$V\$Y\$K\$K\$K\$S\$IF\$V\$LP\$ND\$DL\$K\$L\$A\$R\$K\$G\$T\$R\$E\$V\$P\$Y\$N	1731
Db	1241	EDK\$T\$T\$YK--K\$P\$Y\$S\$R---YPLIS\$N\$F\$IL\$H\$K\$K\$R\$S\$IL\$V\$LP\$R\$E\$LL\$K\$IL\$A\$R\$O\$G\$S\$T\$G\$F\$H	1294
Qy	1732	YN\$K\$P\$AL\$DIP\$Y\$P\$P\$T\$E\$G\$IT\$W\$R\$Y\$L\$Q\$V\$K\$S\$I\$AG\$S\$IL\$M\$L\$R\$M\$A\$S\$B\$W\$D\$M\$A\$K\$V\$P\$G\$G	1791
Db	1295	HAA\$N\$N--T\$IMQ\$C\$S\$R\$P\$L\$F\$T\$C\$M\$S\$Y\$T\$S\$N\$AT\$S\$L\$S\$LA\$Q\$L\$R\$M\$S\$C\$L\$R\$W\$D\$M\$IA\$K--P\$P\$T	1352

[illegible]

```
Db 2353 -LEHHLELKKRSLKRLSLERNLQSEIHE---DVKTQVRHVRPLISNA----- 2397
Qy 2661 TPVLPAPAPAPSPPPPPQVQHTGLSTPTLPY-----ASQKRKREEK--- 2704
Db 2398 -----SPDQSENERGSG---EPRLDPKRTQVONPRHGAQRPKLTTRKKRKLTC 2442
Qy 2705 -----DSSSKSKKKMMISTTSKTRKKTLYCICK 2734
Db 2443 ICTPYDPTKRYVGVCDLCSNMFFHGDVCSITEEASK-KLSEFICIDCKRARETOOLVCSCR 2501
Qy 2735 TPVDESKFYIGDRCQNMWYHGRVGLQSEAEILIDRYVCPQCCSTEDAMTV-LTPITEND 2793
Db 2502 QPDESQSFYICCDKQCDWFGHRCVGLQSEAEFIDRYVCEPCQKNDANANAKKLTSTND 2561
Qy 2794 YEGLEKFLRSLQAHKMAKMPLEFPDNDAPDYGVYIKRPMDLATMEERQORYRREKILTEF 2853
Db 2562 VEBLKLKLINQOMQKSAWPEMEVDPKPADYKYVKEPMDLKRMLKLESNTYTKLSEF 2621
Qy 2854 VADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVQKLKGF 2893
Db 2622 IGDMTKIFDNCRYNPKSESSFYKCAEALSEFVQKIKNFR 2661
```

RESULT 3

```
Q8VDN7 PRELIMINARY: PRT; 645 AA.
ID Q8VDN7
AC Q8VDN7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 72.2 kDa protein (Fragment).
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021489; AAH21489.1; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001965; ZnF-PHD.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 645 AA: 72167 MW: 6486F34F339820F9 CRC64:
```

Query Match 18.1%; Score 2705.5; DB 11; Length 645;
Best Local Similarity 82.9%; Pred. No. 1.2e-100;
Matches 548; Conservative 33; Mismatches 55; Indels 25; Gaps 10;

```
Qy 2251 GAKAKQAPQASARQPOPTQOSPAQAPQVQTPQVQVQVSSHWPSKQPTTHAOSKRP 2310
Db 4 GAENAPQPAQAPAPAPQ--QPPAPAPQVQTOP-----AVSHVPSFETQPSQAOTSKP 55
Qy 2311 QVAAQSPQSNVQSGSPFVQSPQTRIRPSTPSQLSPGQSGVQVQTTQSOPPIQPTSL 2370
Db 56 LVATQCPQSSSVQSGSPFVQSPQTRIRPSTPSQVTPQQPOVQTTASQPIPTPTSL 115
Qy 2371 QIPSGQPSQSPQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSS 2430
Db 116 QAPSGQPSQSPQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSSQVQVSS 169
Qy 2431 VLSQISQVVAQIQAQSGSVPOQIKQLPFIQIQQSSAVQTHQIQNVVTVQAASVQEQLOK 2490
```

```
Db 170 VLSQISQVVAQIQAQSGSVPOQIKQLPFIQIQQSSAAQ--QSVVTVQAASVQEQLOK 226
Qy 2491 VQOLROQOQKKQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2550
Db 227 VQOLROQOQKK-QQLETEFHTLQASNOSEIIOQOVYKKNNAVIEHLQOKKTMTAEER 285
Qy 2551 ENQRMIVQNMVYLLDKIDKEKQAKKRRRESEVQKRSKONATKLSALFKRKEOLR 2610
Db 286 ENQRMIVQNMVYLLDKIDKEKQAKKRRRESEVQKRSKONATKLSALFKRKEOLR 345
Qy 2611 AELTKRALDKLOLQVQBELKRDIKKKEKDLQLAQATVAACPPVTPVLPAP-PA 2669
Db 346 AELTKRALDKLOLQVQBELKRDIKKKEKDLQLAQATVAACPPVTPVLPAPAPA 405
Qy 2670 PPPSP--PPPGVQHGLSTPT--LPVQSOKRRESEKSSSKKKMMISTTSKTRK 2725
Db 406 APPAPRSPPTSHSLPPGHPAPLPVTSQKRRREKND--SKSKKKMMISTTSKTRK 463
Qy 2726 DTKLYCICKTPYDESKFYIGDRCQNMWYHGRVGLQSEAEILIDRYVCPQCCSTEDAMTV 2785
Db 464 DTKLYCICKTPYDESKFYIGDRCQNMWYHGRVGLQSEADLIDRYVCPQCCSTEDAMTV 523
Qy 2786 LTPLEKDYEGLEKRVLSQAHKMAKMPLEFPDNDAPDYGVYIKRPMDLATMEERQORY 2845
Db 524 LTPLEKDYEGLEKRVLSQAHKMAKMPLEFPDNDAPDYGVYIKRPMDLATMEERQORY 583
Qy 2846 YYEKLEFPVADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVQKLKGFKASRSNNKLOST 2905
Db 584 YYEKLEFPVADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVQKLKGFKASRSNNKLOST 643
Qy 2906 A 2906
Db 644 A 644
```

RESULT 4

```
Q9WOTO PRELIMINARY: PRT; 976 AA.
ID Q9WOTO
AC Q9WOTO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CG1135 protein (Fragment).
GN E(BX) OR CG1022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY.
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Neilson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
```

RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Klamm B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mishina N.V., Moberly C., Morris J., Moshfeghi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.W., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gdbbs R.A., Myers E.W., Rubin G.M., Venter J.C. ;
 RT "The genome sequence of *Drosophila melanogaster*." ;
 RL Science 287:2185-2195(2000) .
 DR EMBL: AF003467 ; AAF47363.1 ; -
 DR FLYBase: FBgn00035118 ; E(ex) .
 DR InterPro: IPR0000637 ; AT_hnck .
 DR InterPro: IPR004022 ; DDT_dom .
 DR InterPro: IPR001965 ; ZnF_PHD .
 DR Pfam: PF02791 ; DDT_1 .
 DR Pfam: PF00628 ; PHD_1 .
 DR PRINTS: PR00929 ; ATHOOK .
 DR SMART: SM00384 ; AT_hook ; 1 .
 DR SMART: SM00249 ; PHD_1 .
 DR SMART: SM00249 ; PHD_1 .
 SQ NON_TER 976 976 6A6C76E00DCA768 CRC64 ;
 SEQUENCE 976 AA 112510 MW;

Db	534	-----	533
Qy	599	HTPVSIOETVCGFTSEKSTGSELSESPGAGKGSAGSTRIITRLRPDSDKLSQOYAA	658
Db	534	-----	547
Qy	659	AAHAAKLFKEGKEVLYVNSOGELISRLSTYKKEVIMKGNINNYFKLGOGKRVYVNOYST	718
Db	548	-----	570
Qy	719	NSFALNKHQHRHEDHKRRHLAHKFCUPLPAGEFKWNGSVHGSKVULTISTRLTLTOLENNI	778
Db	571	NPLANKPQRNEEDKRRHLSHKFLSTLSDAFKMGITWGTDNMITTLRQGLIFESNI	630
Qy	779	PSSEFPHPWASHRAHNTKAVQMSCKPREPRLALALIECAVKRYVMPLPIREFLGTRHR	838
Db	631	AASTLINTWVYVKKIKINWAAVNNARSPSEFVYVLLTQASLSKVPANWHDGLHTTLQR	690
Qy	839	MTSILEREKEKVKYKKEKKO---EEETMOATVKTTFPVKHOVWKKOGKEEYRVYGGM	895
Db	691	ITSERERKFLERKEKREDDDEERRLAFNIKTLTGKIQWKKOGKEEYRVYGGM	750
Qy	896	SMIST-----HYVRFPKPLPGNTVNYRKSLSEGTNNMDENMDESDRKRSRSPKIKI	950
Db	751	LWLSSRRCGVBARAQPLTHNRYVHYTM---GEENDVNEIT-----LVDPRTQRF	799
Qy	951	EPDSEKDEYKGSDAKADQ--NEMDSKTTEKKDDVVELDSDSKRKEPEVEDD	1008
Db	800	MOQCESSNVDOCVHYLPDYQKVKYEDTE-----	831
Qy	1009	MKTESHVNCQESSOVDPVNVSEGHIL--RSTYKKTSSKDLGLERRIKOFTLEKQRL	1066
Db	832	KING-----IDVSKALNAGRGTYYSKVARKSRLDDLDRRLKLAIVEEQM--	877
Qy	1067	EKIEGGIGKIGTSTNSSKNLSSES	1092
Db	878	ASKIPSDMKPLLVSONNTANSKQT	902
RESULT 5			
Q9H5E0			
AC	Q9H5E0	PRELIMINARY;	PRT; 412 AA.
DT	01-MAR-2001 (TREMBlrel, 16, Created)		
DT	01-MAR-2001 (TREMBlrel, 16, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel, 21, Last annotation update)		
DE	CDNA: FLJ23531 fis, clone LNC00665.		
DS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISUE=LUNG;		
RA	Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,		
RA	Kawakima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,		
RA	Yanada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,		
RA	Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,		
RA	Isozaki T., Sugano S.,		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK027184; BAB15686.1; .		
DR	PRINTS; PRO1582; KV33CHANNEL.		
SO	SEQUENCE 412 AA; 46539 MW; D726D830BB12B94 CRC64;		

QY 223 K-----VLQFLVDOFLTNTIARELMEGVIOVDHCRVCHK-IGDLIC 265
Db 375 QDESFEPFNKHFRLVYLFPMSYRFLLYSSEFKLVNNVNGFQDNENCRVSGSSGRVVG 434
QY 266 CETOSAVYHLBCVPRPLEBEDEMOCEVAAHK-VPGVTCOVAIQOKNKPRIHREPIGY 324
Db 435 CTQCEAAHFVBC--SHLPFPE-VLVCNICKKNSAVRGVLPPEDEVDPB-LRQOPIGR 489
QY 325 DSRKKYWFNLNRLLIEEDTENENEKKIMYVSTKVOLAELIDCLDKDWEAELKILEEM 384
Db 490 DRYGTYVFIYRLVQSLDETE---LYYSTVPQLVQLQKDRYAYEKOLCOTIRLR 545
QY 365 REELHRHDITEDLTNKRGSNKSFLAANEFILESTIRAKK---GDIDNVSPREETDK 441
Db 546 IDEFLQOMALIVEMTSERR-----EALLETWVKROLIGYFAEATPQ----- 588
QY 442 NETEDSDAKENREFPDQSLKSDOKTPD---DDEPGCSEPTVEGDKGNSVANL 498
Db 589 -----IYLHRDSMKRMASILRDCAKQGVKOEVL----- 618
QY 499 GDNNTNATSEETSPSEGRSPVCTSETPDSSNMAEKVASELPDVPPEPNKTCESNNTS 558
Db 619 -----EE--PYEGOSPVKCVQFVEDSI-----LPE----- 641
QY 559 ATTTISIQNLENSNSSSELNSSQESAKAADDPENGERBSHTPVSIQOEIYGDFTSEKST 618
Db 642 -----SMIGIFPAKLINTFWSGA-----TOEELVQEFV---DI 672
QY 619 GELSPGAGAGASGSHIITRLRNPDKLSQLKSOQVAAAHAANKLFKECKEVLVYNS 678
Db 673 SDNFPAPSA----- 681
QY 679 QGELISRLSTKKEVINMGNIINNYFKLGQEGK---YRVYHNOYSTNSFALNKHQREDHDKR 735
Db 682 -----NLMRMGDEGNQDTEMTYNYYSRNMSESEFLTRKKAADKK 721
QY 736 RHLAKFCLTPAGEFKW---NGSVHGSKVLTISTRLTLTLOLENNIPSSFFHPWASHR 791
Db 722 KYMSKF--AQIDNDWVYAKNROFYGDASLCKEFIMMTLOQVININIPDLHRRKPEFA 779
QY 792 ANWIAVOMCSPREFALALILECAVVRVYVLPIMREFIGTRLHRMTSTEREKEKYK 851
Db 780 KGFDELEVADYKLVCLKLKLDCAVKTITFPMQWNGLGOTRLERTIVDQREPNMEQ 839
QY 852 KKEKK-----QEEETMOQATWVKTYPVVKHQVOKQGEERYVTG---YQWMSMISKT 901
Db 840 QRLKIIDADALTKDLDSEFVRVNYMKPKWPNY-ILRQGEYTRNAGSGMGMVAAK 898
QY 902 HYRFVPLKPGNTVNVYKRSLEGTANNMDEMDSDKRCSSPKKIKIEPDSEKDEVKG 961
Db 899 YVEKMI-----QVPESPKPLPLAVTVEIKTESVSNRK--- 930
QY 962 SDAAGADONEMDISKITEKKDQDVKELLDSDDKPCKEERPEVDDMKTESHVACQESS 1021
Db 931 -----ARLELVLKIT- 942
QY 1022 QVDVYVNSEGHILFTSYKKTSSKLDGLLEBRIKOFTLEEKORLEKIKLEGIGIGKT 1081
Db 943 -----KKRQSG-----GKS 952
QY 1082 STNSKNLSESPVITAKEGCOSDMROQSPNANNQPEDLIQCSOSDSSVLRKMSDPS 1141
Db 953 SKKPFELT-----NGCYSFSCRS----- 971
QY 1142 HTNKLKYKRDVLDVYSIRSETPKCPKONSTIENDIEEKVSDLASRGQEPKTKTGNDFF 1201
Db 972 -----NPNRKC-----YSP----- 980
QY 1202 IDDSKLASADIGTLICKNKKRKLQIESDQIVSSSSKALHSSVFKSTNDRCATPLSRAMD 1261
Db 981 -----MCNN----- 984
QY 1262 FEGKIGCSESNSSTLSSSDTVS:IODSSEEDMIYONSNEISISEOFRTREQDVELEPIKC 1321

Db 985 ----- 984
QY 1322 ELVSGESTGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSQJIKLKNTTDDKNNENRE 1381
Db 985 -----GYLSAKQAHDERKLEB----- 1001
QY 1382 SEKKQRTSTQINGKDNKPKIYKGECLKEISESRVYSGVNEPVNNINKIIPENDIKS 1441
Db 1002 ----- 1001
QY 1442 LTVKSAIRPFINGDVIMEDFERNSSETKSHLSSDAEGYRDSLETLPSTKSDSTQ 1501
Db 1002 ----- 1001
QY 1502 TTTPASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLSDNFIDENGLPINKNEN 1561
Db 1002 -----SG 1003
QY 1562 VNGESKRKTVITEVTMTSTVATESKTVIYKKGDKQTVVSTENCASVTYTTTTVTYK 1621
Db 1004 VLGEKK----- 1009
QY 1622 LSTPSTGGSVDIISVEKQSKTVYTTVDSLTTGGTLVTSMTVSKEYSTRDKVYLMKFS 1681
Db 1010 ----- 1009
QY 1682 RPKTRSGTALPSYKREYTKSTKSTKSTFVLPNDLKLARKGIREV--PYFNYNAKPALD 1739
Db 1010 -----AMPREIOTFTSTRKGSIFVLQKILKQIMMGCCQOYVMPFGSAGIKSNLL 1062
QY 1740 IMPYSPRPFQITWRYRLQTVKSLAGVSLMLRLMASLRMDMAKVPYGGGSTRTETS 1799
Db 1063 IMPYAPRPTLDCMKQWQJLNARSLHAYVALQKTIIMSSIKFNEFD--PDDHPPRRVYI 1119
QY 1800 ET--EITTEITKRDVGPYIGRFEYCIRKIIICPI-CVPEPTKEPPTQORQK-----LR 1850
Db 1120 DTPSHDERRRRIIRHKEMPYGQYEREMEITIPLYDEPEEDESGLSNRRGSSSEFSHR 1179
QY 1851 SALSRLPRET-----PKOT-----GVIIETVVAEBELTME----- 1883
Db 1180 SSARKKRPQORHEFLSLKFGNTPKKSNAFRSLDNRRAIRREWDGVTLLKVPKIDYK 1239
QY 1884 --IRAFAERVEKEKAQAV-----EQAKKRLQOKPTVIATSTTSPTSTSTISPAQ 1934
Db 1240 KIRAEAKTARKLEATRKKAQAKADEERRRIQQOQOSVARIRV-PMHS-----LIPSE 1294
QY 1935 KYVAPAPISGV---TTGTKMYLTTKVGSPATVFOQ-----KNFQGTATVYKQO 1983
Db 1295 RNNV-PYLSQOQORRPNRNGERGFLEKYNNSSSVSDQAHGYASTPPGYHQPOPNIRQAG 1353
QY 1984 SNSGVVQVQOKVLGIIPSSGTSGQTFSGFORPTATVTRPMTSGGTSOYITGQ 2043
Db 1353 YNQ-----LPKRPSTSPNFOS-RP-VATIPPTPLRAAGADG--VVRAYM 1396
QY 2044 IRPG---MTVIRPLOSSTLGKAIIRPVVNVQAPAPQVMTQIIRGO----- 2087
Db 1397 MTPGNKSTYNTNSTPYPO-----ALNRQOYQDROQOQAPAVARLNGYHFMGTGRGGR 1451
QY 2088 -----PVSTAVSAPNTVSTPGOKSLTSAT-----STSNIOSS 2120
Db 1452 NPSYOMHQLPQNRALQRFGESESTTEMRYTEAALPDNDGDEQPPVJPRXDPTSNFPAQ 1511
QY 2121 ASQPRPQOQVYKLMMAQLOLTQGHGSGOGGLVYIQGGGQTTGLOL-----IPQGVYVL 2176
Db 1512 RAQOQHPOSRRPVSTPAQIRTTQPGVKN--NVLKMASDQTKMVLPKPGQFPPTGYI- 1568
QY 2177 PPGQOLMAAMPNGTVORFLTPTATTATTTATTTVSTTAAGTGEORSKL----- 2230
Db 1569 -STGQRYVYRQPTAVQQROLYT-----ATPGTIVVIRPNANGAPRQODHQVBRVYQ 1621
QY 2231 --SPQ-MOVHODKTLRP-----AOSSSVGPAPAKQPOTAOPSARPO-----PQTQ 2271


```

Db 1622 ASGPRAMEYNDQGTPEPGQVRYVLOGSGSTPNVNPPEKVSRRGPRGLTMMOVQOQ 1681
QY 2272 POSPAQ-----PEVQTPREVQOTVSSHVPSDAQ 2301
Db 1682 QHNPFOAHYMDPDDATGFAVSTTT---EQVPDEOQ 1713

RESULT 8
0960Y3
ID 0960Y3 PRELIMINARY; PRT; 1022 AA.
AC 0960Y3:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE ID30146P.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY051776; AAK93200.1;
DR Flybase; FBgn0035118; E(BX);
SQ SEQUENCE 1022 AA; 111667 MW; 1174939B61962E63 CRC64;

Query Match 4.6%; Score 692.5; DB 5; Length 1022;
Best Local Similarity 25.3%; Pred. No. 5e-20;
Matches 294; Conservative 157; Mismatches 440; Indels 269; Gaps 47;

QY 1520 DMEIETSEVKYKTSPTSEESNL-----SNPFID---ENGLPIK 1558
Db 6 NVEIESTEDSIYTLGNSGAEVDYDTPGRRKRNOKSKSYIGTKRYVLDQTLDKDIPLAK 65
QY 1559 NENVNGESRKRTVTEVTMTSTVATESKTVIKVEKDKQTVVSTENCASKVTITTTT 1618
Db 66 -----QNRREPIT-----ARPKREC-----VKYERETENGNERV----- 97
QY 1619 VTKLSPSTGGSV---DIISVKEQSKVYTTVYDLSLTGGTLVTSMTVSKYSTROK 1674
Db 98 ---YSTSSPRGRVYLLINDAKLYEQA-----VKTEDEK 126
QY 1675 VKLMKFSRKRTSGTALPSYRKFEVTKSKSLFVLPNDLKLARKGIREVPYFNVA 1734
Db 127 SRTTK--KTSYSR---YLLISFLTHKKKRLSLVPRRELKLAKGSSKSTNGHHNA 180
QY 1735 KPALDIWYPPSPRPTFGITWRRLQTVKSLAGVSLMLRLMASLRMDMAAKVPGGGST 1794
Db 181 KNN-TWQOCSPLEFRTQWSYRTSNATSLASLQLRITWISGLRMDMDIAK--PSTDK 238
QY 1795 RFEISTEITTEIIRRDVGPYGIREFYCIRKIICPIVPEPKTPPORKGSSSL 1854
Db 239 HOYTTEITVTELLKRLKRSKRGKSTYLRKRVIPLEPKVREV--TSISGLR----- 293
QY 1855 RPKRPETPRQGTGVILETVWAEELWEIRAFEREREKQAQVDAQ--AKRLEQOQPT 1913
Db 294 KKKRAASPTEPQITEWVDEKLEMEIKFMGEQEKARLSAVTRVSAHQLE----- 348
QY 1914 VVATSTSTSTSTSTISPAQKVAAP-IGSVTTGTKNVLTLYGSSPATVTFQOKNFH 1972
Db 349 --ASGNGSNTSTNGALGAAGROQLAPKLEEDVK--EKMEQOLKL--ORAVHQQRK--- 398
QY 1973 QTFATVVKQGSNSGVVQOQVLS-----IIPSTGTS---QOTFISQ-----PRATVY 2021
Db 399 -----LVATGELTRSVTPVKGOVIGSRRIYVKNPDGTRIIQOAVVQVSRTGANTAAA 453

```

```

QY 2022 TRPNTSGSGTTSNS-----OVTGPQ-----IRPGMTVIRTP-----LQOSTL 2060
Db 454 ASPTVGGSTSTOSNFBSTPHKVOIIRKPDGKVSXGLNPGQOLVMPGKRLHVLTTTTS 513
QY 2061 GKAIIRTPVWVPGAPQOVMTQIIRGQPVSTAVSAPNTVSPGOKSLTSTSTNSIOS 2120
Db 514 SNSAGGNKMKVPIKPASTSS-----SPAISSAQTTNTPVTPIKQIAVHVTKNSATQSI 569
QY 2121 ASQPRPPQOQVKLMAQLTQLTQHGNGOGLTVYIQSGGQTTGQ-----LQLIPOG 2172
Db 570 AS-----SSRYALPLAQI-----KNKLLAQOQOQSTSSPATSSPVKIVSK 613
QY 2173 VTVLPQGOQLMAAMPNTVORFLETPPLATTATTAATTTTIVSTTAA---GTGEORQSK 2229
Db 614 VVNTSTSGQTLQOVVQSSS-----KLTVGQNAQOGKVIITSTMAQOQSTSVYQOQ 665
QY 2230 LSPOMVHQDKTLPPAOSSS-----VGPAK--AOPRTAQPAPQOPQOPSPAQPEVO 2281
Db 666 LVQSPFIQSS-----PQISMTOQOQIVVGGQRIILSPGQITVYQRNVVPGQALQMVQOQIQ 721
QY 2282 TQP-----EVQOTVTVSSHVPSAOPTHAOSKRPYAAQSOQSNVQSGSPVRYQSP 2333
Db 722 TQOQOQHNVQOPQOQEVVQSNQIVQSSPSAQTKLVKQLVWQOQSOQTIEKTTQTTTDS 781
QY 2334 SQT-----RIRSPQSLSPGQOQOVQTTTSPPIPIQPTSLQIPSOQPOSOQPOVQSS 2388
Db 782 NETGQOVVLPVPSLTAQQLAGKQLQVATVNGQOVIK-----LGNNQADLVNHI 831
QY 2389 QTLSSGQTLNVQSVSSPSRPOQLQI-QOPQOYIAVQLOQOQOV--LSQIQSOVVAQIQAO 2446
Db 832 KIQGDSNAHIVTNSNATAVPAQNPQSPVVKQOALPQSPQOVVQVQOQIHOQSPNTEFSG 891
QY 2447 QSGVPOQIKL-----QPLQIQSSAVQTHQION-----VTVQASVQ----- 2485
Db 892 VPIPTQOQVLTQAVQAPAP-QQALSVESLQNOQPPGYIKCVIQAQVLOTETGPRIVLOG 950
QY 2486 -----BOLQRVQ-QLRDQOQKKKQOQIEINVTPTSKLIRKELIQOVYMKHNAVIE 2536
Db 951 LVGNDFTAQQLQVQVQVQOQLMKAQESNGKLGIVGPTIYIAVQ-----PENAV-- 1000
QY 2537 HLKQKSMTPAREENQRM 2556
Db 1001 -QOQPELTPVHQSAHQOV 1019

RESULT 9
Q9W0T1
ID Q9W0T1 PRELIMINARY; PRT; 997 AA.
AC Q9W0T1:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG7022 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Padlos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kohira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskehn D.R., Pacled J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL ENBL; AE003467; AAF47362.1; -
DR FLYbase; FBgn0035118; E(bx); -
SQ SEQUENCE 997 AA; 109295 MW; 8106E06CF6A8C7 CRC64;

Query Match 4.4%; Score 661.5; DB 5; Length 997;
Best Local Similarity 25.4%; Pred. No. 8.4e-19;
Matches 286; Conservative 151; Mismatches 409; Indels 279; Gaps 48;

QY 1520 DMEIETSEVKKVTSSPTSEESNL-----SNDFTD-----ENGLPINK 1558
D 6 NVEIESTEDSLVTGLNSGNAEDVMTPEGRKRNKSKSVYIGTGDVLDQTLDKRIPINK 65
QY 1559 NENNVGSEKRTVTEVTMTSTVATESKTVIKVEKGDKQTVSSTENCAKSTVTTTTT 1618
D 66 -----ONRRFPIT-----ARPKREC-----YKVERETFEENGNERV----- 97
QY 1619 VTKLSTPSTGGSV-----DIIVKROSKTVVTTTVDLSLTGGTLVTMTYSKESTDK 1674
D 98 ---YSTSSPRGRVLLNDAAKLYRDA-----VKTEBK 126
QY 1675 VKLKKFSRPRKTRGTALPSYRKFTKSKSIIFVLPNDLKLARKGIREVPYFNA 1734
D 127 STIRK-KPVSRS---YPLISNFLTHKKRSLVLPREFELKLARLGKSTNGFHMAA 180
QY 1735 KPALDIPYSPRPTGTTMKYRIQYKSLAGVSLMLRLNASLRWDMAKVRPGGSGT 1794
D 181 KNN-TIMQYQCSRLPFTCMYSRTSNATSLSLALQILRLMSCLFMDMIK-PPSTGK 238
QY 1795 RTESETETITTEIKRRDVPYGRPEYCIKTIICPGVETPEPTPPKRGIRSSAL 1854
D 239 HOVTDTIEITVLELKLHSGRTGKESYLRKVVIPLEMKYIAEV-TSTRSGR----- 293
QY 1855 RPKRPETPKQGTPIIETWVAEELEIRAFERVEKEKAQVEO-AKKRLQOQKPT 1913
D 294 KRRAESHPQPEQITEWVEDKLELMEIKFEGKEKAKLSAVTRVARSROLE----- 348
QY 1914 VIATSTSPSTSTSTISPAQKWAAP-IGSVYTTGKAVLTITVGSFATVTFQONKPFH 1972
D 349 --ASGSNSNSTNGALGVAGVOLAPKLISEDVK--EKMEQOLKL---QRAVHOQRK--- 398
QY 1973 QTFAFWKQSGNSGVVOYQOKVLG---ITPSNGMS---QGFETSQ---PRTAVT 2021
D 399 -----LVATGETITSVTPVKGVIGSRVYIKNPDGTTIRITIQAVTQVSRTGANTAA 453

QY 2022 IRPNTSGSGGTTSSNS-----QVITGPQ-----IRPGMTVIRTPLOQSTLCAIIR 2066
D 454 ASPTVGGSTSTQSNBSTSTPHKVIIRIBDGVSVRGINPGQQLVQMP-----DGKLIHL 508
QY 2067 TPVWVQPGAPQ-----QVMTQI-----IRGPVST-----ANSANVSS 2101
D 509 TTTTSSNAGOGGLVFKLKSNTITWVYHFLGKMKVPKIPASTSSPALSSAQTTN 568
QY 2102 --TP-----GOKLSATSTSIQSSASOPRPPQGOYKLFMAQTOLTOGHGNOGLTV 2154
D 569 PTPVPIKQALAKVHYKNSATQSIASS-----RALPLAQI-----KNLL 609
QY 2155 VIQGGQGTGQ-----LQILPQGVTVLPQGOQLQMAAMPNGTVORFLETPLATAT 2206
D 610 LAQOQOQSTSSPATSSPVQKIVSKVNTSTSGOTLQOVFVQSGS-----KLTVGQ 661
QY 2207 TASTTTTIVSTTA---GTGEOROSKLSPOKOVHD-KTLPPAGSSVGPAP--AQOTA 2260
D 662 NAQGGKVTISTMAQOQSTSPVQOQQLVQSPPIQSPQOISMVTOIIVGCGRIILSPQT 721
QY 2261 QPSARPOPTOPQSPAQPEVOTP-----EVOTPTVSSHVSEAOPTHAQSSKPOV 2312
D 722 IYTGNNVQSQALQVNVQOQIQOQOQOQHVVYQOQQPVVQSNQIVQSSPSAQTLVQL 781
QY 2313 AAQSQPQSNVQGSFVRVQSPSQT-----RIRSPQSLSPEQSOQVQYTTTSQPIPIQPH 2367
D 782 VYQOQSQGTIEKQITTTDSNETGTQVLVNSLQALQAGKQLQVATVNGQYIVP- 840
QY 2368 TSLQIPSOQPOQPOVQSSSTQTLSSGOTLNAVSVSFSRQLOIQ-QPQVIAVPLQ 2426
D 841 -----LGNQQAQIYAHIKHOGDGNNAHTVTSNATVAPQANQTSVQKQALPQSP 891
QY 2427 QOVQV-LSQIOQVYVAQLOAQOQSGVPOQIKL---QLPIQIOSSAVQTHQIN----- 2475
D 892 QOVVYQOQOIHQOQSTNEBESGVTPIITQPVQLQAVQAPQ-QQALSVESLQONPQTV 950
QY 2476 --VTVQQAASVQ-----EQLQRYV-QLRDQOK 2500.
D 951 IKCVTAQVLQFHEGPRIVLQGLVGNDFTAQQLQVLQTVQKQOLMK 995

RESULT 10

QY 09VC00 PRELIMINARY; PRT; 2768 AA.
AC 09VC00:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG13648 protein.
GN CG13648.
OS Drosophila melanogaster (fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Pterigolidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle G., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Padlos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,


```

Db 1470 STEPSAEVEKASGETSESNDNEIDAGASSTPVPVSADEKTPSTKTEVTE---ADDKFTTV 1526
Oy 1534 SPIT--SEESNN---LSNDFIDENGLPI-----NKNENVNGESK-----RKT 1570
Db 1527 APLAGEEESNLPKPLQDIFEEBA--PVAATTAAPSKDD---GQOKVEVEBEKIEDGQRP 1582
Oy 1571 VTEVTMTSTVATESKVIKVEGDKQTVVSTENCASVTVTITVTYKLSPTSGS 1630
Db 1583 IEDETSTPTS-----SENEIEPESDRATTAIPASKEE--PSEPSTGAPTKDEPAEPSTADP 1635
Oy 1631 VDIISVKEBSKVYTTVYDLSLTGTLVSNITVSKVSTRKVLKMFSRKKTIRSQT 1690
Db 1636 ESDSE--KETPESEVPTTAPA---GEKIPSTSTPDEEPTATSAVAPKDEDEVEKETST 1690
Oy 1691 ALPSYRKFTYKSTKKSIFVLPNDDLKKLARKGIREVPYFNNAKALDIWPPSRPTP 1750
Db 1691 EIPTOAPASSEEDENS-----STDQIPS-----EVP-----EKKP----- 1720
Oy 1751 GITWRYRIQTVXSLAGVSLMLRLMLASLMDMAKVPYPPG---GSTRETSETETITTE 1807
Db 1721 -----ETPAQTPEBGDIVGATPAPTSDVEYPPVQ 1749
Oy 1808 IIKRNDVGYGIREFYCIKIKIICPIGIVPETPKTPPTPQK--GLRSALKRKRPETKQIG 1866
Db 1750 RLPE-----EVLAEIIPQSTETGIRKODETQAAPSIDRK-- 1783
Oy 1867 PVIETVAEELELMEIRAFARVERK--EKAQAVEO-----QAKRLRLEQOKPTVIATST 1919
Db 1784 -----EYVTEIDEATVAPISKEDEKPTPEBEKPVQOKPTGEPSEBEKEKIEDOVST 1839
Oy 1920 TSPTSTSTTISPAOKVMAPISGSVTTGTKVLTTRKVGSPATVFOQKNFHQETATWV 1979
Db 1840 EBPVSTFEASEAGSTES-----SEEVKPTSEGVAKKPED----- 1873
Oy 1980 KOGQSNM--GVVOVOQVVLGLIBSSTG---TSQOTFSFQPRKATVITR--PNTSGSGCTTS 2034
Db 1874 KQPSSTAQAAPVETIPEIISTELPQADQDKPTSEKAVPVDSEDTSADESKIPSVSG----- 1927
Oy 2035 NSQVITGPQIRPQMTVIRTPLOQSTGKALIRTPVWOPGAPQOV--MTQIIRQOPSTAV 2093
Db 1928 --EVEGPEV-----TTASQAABEDELKTPASEPSTKVKVETEVOKFEDETKAD 1977
Oy 2094 SAPNTVSTPGOKSLTSATST-----SNIQSASQPPRPQOGVKLTMAQLTQLTGCHG 2148
Db 1978 EMPESVTQVSDVATSTSPAPAGGDIKDEQATYASP--EBEIEIKPTIAPAAEIPQ----- 2031
Oy 2149 NGLITVIGQGGTGOLOLIPGVVLP--GFGQOLMG--AAMPNGTVQRFLETPLATTA 2205
Db 2032 -----PSEKEPVDE--QEVESGTATPAESDQDPIDEIAPATSGPIDE-----ASTA 2076
Oy 2206 TTASTTTTVAAGTGAQROSKLSPOQOVHODK-----TLPPAOSSSVCPAKAPOPT 2259
Db 2077 APTKESTTIVASGA-----SP--AVHDEIKDVTITTOPVADKEKVAAPQDETKT 2123
Oy 2260 A-----QPSAR--POPOTQPSAPQEPVOTQPEVOT--QTTVSSHVPs---EAOPTH-- 2304
Db 2124 SIDVSTDSPTAQDDEKQDTEAPVATVYSSPTADSAASSTPTVAVPSVEIDTQPMMD 2183
Oy 2305 --AQSSKPOVA--AQSPQSNVQGSQPVVR--QSPSQTRIRPS-----TPSOLSPGOQSQ 2353
Db 2184 IMSQTTAAPTADCAASTFEDQADAPVTPSPDAEKTVPSPAPQSDKTPSSAPADDADE 2243
Oy 2354 VQTTTSQ-----PIPIQHTSLQIPSGOPOPOVOSTQTL-----SSGQTLNVS 2401
Db 2244 IRATATPLDDNKIRATVAACQTDGVPATAPALDEDKIQTAAALDEKIPSTAPALDDEK 2303
Oy 2402 VSSPSRQOLQOPQOVIAVQLOQOVYLSQIQSQVVAQIQAOQSGVPOQIKLOLPQ 2461
Db 2304 IPAPVSPVVFDEVPSESEKPAVSEYDE-----ESTEPVHVDVETSTDEPTSDAKLRPMS 2358
Oy 2462 IOQSSAVQTHQIQNVYVQA-----SVQOLORVQOQLDQOQKKQOQIIVINVPNSKL 2516
Db 2359 APATPESPATEAIVETTAAPALEKEVPEKATEQPELEKETPEKATEQPELEKETPEKA 2418

```

```

Oy 2517 LIKVEI-----IOQVNMKNHNAVIEHLKOKKSMTPAER-----EENORMIV 2557
Db 2419 TQOPELEKETPEKATEQPELEKEVETKATQEPESVDEKTPPEVVAPPSLDSTDEDE----- 2474
Oy 2558 CNOVMKYIIDKIDKEKQAAKKRRREESVQOKRSKONATKLSL----- 2601
Db 2475 -----ESVSEESADKKKDKNKETEEDTDKKEEPEVAVVEIQPQSEAAVPTTG 2525
Oy 2602 --LEFKHEQOLRAELIKRLALDDLDQIEVQEEKRLKRIKKENDLQOLQATVAAPCP 2659
Db 2526 HPLFPH---TASSTTPPAVDDR-----VGE-----DEENTYKLSSTTTSTTESP 2570
Oy 2660 VT-----PVLPAAPPAPPPSPPGVOHTGLSTPTLPVASQKKRREEDKSSSKKKKM 2715
Db 2571 VTSAPSTTVYASQOQOPPIPTPPYG--HA-----PEYDEIDEVEVPGTCRYAGKL 2620
Oy 2716 ISTTSKETKDKMLYCICTKTPYDESKFYIGCDRCQNMWYHRCVGIQS-----EAL 2767
Db 2621 YVSAQOIPRDDPCDFCFC-----FRSDIICLOQSCPPPIAGCHEEP 2661
Oy 2768 IDEYVCPQO-----STEDAMTVLRP 2788
Db 2662 ISGFCPCRYECPVMAAVLNTTSTTTTSTTLPP 2695

RESULT 11
Oy 09NM17 PRELIMINARY; PRT; 5476 AA.
Ac 09NM17;
Dt 01-OCT-2000 (Tremblrel. 15, Created)
Dt 01-OCT-2000 (Tremblrel. 15, Last sequence update)
Dt 01-JUN-2002 (Tremblrel. 21, Last annotation update)
De Split ends
Gn SPEN OR CG18497.
Os Drosophila melanogaster (Fruit fly).
Oc Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Oc Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Oc Ephydroidea; Drosophilidae; Drosophila.
Ox NCBI_TaxID=7227;
Rn 11)
Rp SEQUENCE FROM N.A.
Rl MEDLINE=20157049; PubMed=10655223;
Ra Reby I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,
Ra Suh C., Voas M., Williams A., Rubin G.M.;
Rt "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
Rt Protein Kinase Signaling Pathway That Interact With the yan Gene of
Rt Drosophila Identifies Split ends, a New RNA Recognition Motif-
Rt Containing Protein."
Rl Genetics 154:695-712(2000).
Rl EMBL: AF184612; AAF26299.1;
Rt HSSP: P09651; 1HA1.
Dr FlyBase: FBgn016877; spen.
Dr InterPro: IPR000504; RNA_rec_mot.
Dr Pfam: PF00076; rrm; 3.
Dr SMART: SM00360; RRM; 3.
Dr PROSITE: PS50102; RRM; 3.
So SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match 4.1%; Score 613.5; DB 5; Length 5476;
Best local similarity 19.5%; Pred. No. 5,1e-16;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;

Oy 305 DCAAELOKKNPYIRHHPICIGDRSRKRYWPLNR-----RLIIEDOT 344
Db 1855 DTADKAEKKNQ--RHBEKEREKEREKEDLRKQVEEREKDRKAQOEEREKDRKAREEK 1911
Oy 345 ENENEKKIYYSTKYVOLAELIDLDKDYWEAECLKILEEMREIHHMDITE-DLTNKR 403
Db 1912 EREREKKAQEDRK-----KEREREKLR--KEDRKQEKKEKELREKDLREKQ 1958
Oy 404 GSNKSFLLAANEELISIRAKKGDIDNVASPEETEKDKNETENDSKDAEKNREEDQSL 463

```

Db 1959 -----RERDNREKELRDKDLREKEMREKQREKLEHREKQDREHNR 2000
QY 464 EKSDOKTPDDDE--QKSEEPTE-----VGDKGNSVANLGDNTNATSEETSPS 513
Db 2001 EKESRRADYEDQBGKRRRELSYOKSKMDIAGEASSLTAIDCCHNKFNADTLTIOGT 2060
QY 514 EGRSPVCLSETPDSSNMAEKYASLEPODVPEPNK--TCESSNTSA-----TTST 564
Db 2061 PGASP-----STP--SDNTPKERSRKLRSNRPVLRHKLRLSIOESNHSAGGSGGSSH 2113
QY 565 QPULENSNSSELNQSSEAKAADPENGERES--HTPVSIOEIVGDTSEKSGELS 622
Db 2114 QIHHEVDYKRIREMSONISVHSSNORLNDNRDSKEHKSSEFKED-----KNSSGHIS 2166
QY 623 ESPAGGAGSGSTRITRLNPNDSKLSQLKGOVAAAHANKLFKCKGEVLVANSOGEI 682
Db 2167 RPHGCGSSSSSKHHNR-----RDKHHQKGSASITNSSI-----EYVVDPISTQK 2214
QY 683 SRLSTKKEVIMKGNINNYFLQEGKYRVYHNOYSTNSFALNK-----HQHR----- 729
Db 2215 HNLTSEELQSHQ-----PKREKREHPSHANSSSRHKSXRDHHRREKRRHSA 2267
QY 730 -----EDHCKRRLHAKFLT---PAGEFKNGSVHSGKVLITSLRLTLTIOLENNIP 779
Db 2268 ESTWTDEHPPQOHNPRRIRISAAGSGAGELSSAATWTSCKLHHQHRRSVERKSSRGS 2327
QY 780 SSFPHPMASHRAWIKAYOMCKPREFALALALECAVPVVALPIMRBFLGHTRLHRM 839
Db 2328 DEGHSSSKSLRAKMLMSADSDTDDA-----SKHSTI 2362
QY 840 TSIERE-----EKERYKKEKKEEETMOATWKYTFPVKHQVMKQKEEYKVTGYG 894
Db 2363 FDIPODCPNVSMYDKVARCKNMROAEKIKAKPS-----QKQSRAKKKTSTSYD 2417
QY 895 WS---WLSKTHVYFVYKLGNTVWVRKSLBEGKNNMDENMDESDKRSRKKIKIE 951
Db 2418 DSDTEFDROH-----RNSGSSSFHGR--YPLGSSSDDDDETHORRIS----- 2460
QY 952 PDSEKDEVKSDAKGADONE-----MDISKITEKKDOVKLELDSODPKCKEPEMYD 1006
Db 2461 -----SBSDAFHGODNOGASTLADANRVROMO--QNLRLCDGDS---SED--EIR 2506
QY 1007 DDMKTESHVNCQESSQYDVVNVNSEG-----PHLRTSY-----KKTKSSKLDGL 1051
Db 2507 RNMYKHHFGRKNSNSTRIASDSESOQAPADLTIKOEHHIAPAQOEIKRQOLSDFEQFK 2566
QY 1052 ERRIKQITLEKOKLEKIXILEGKIGKIGKSTNSKSKLSPVTTKAKBEGQDSMGOEQ 1111
Db 2567 SRHDSNSISIER--KLTETREI---KTELGDFTNSSEYTYGKLEKXSPETRRKHKK 2618
QY 1112 SPNANNOQPEDLIQGSQSOSDS-----VLRMSDPS---HTTNKLYPKRVLDDVSI-- 1159
Db 2619 SKRR-----LKSSSTADTSAOQPLVMTPLRPSIFDVHSSSECKTKCDNPNDLTEC 2670
QY 1160 -----RSPETKCPKONSINDIEEKYSDLASRQOEPTKSKTYGDNFIDSK 1206
Db 2671 SSIPLEISAGERRRKHKEKKEKREKLRNMTAEATVPN-----SEPTTNDT-----SSE 2716
QY 1207 LASADIDGITLCKNKKRKLIOESPTIVSSKSALHSSVPMSTNDORDTPLSR----- 1288
Db 2717 KLSKEEHRRLKSKSKSMONSCNTKIYNSGA-H---PSTSPSLPAPTTPSAPSTAOTSK 2772
QY 1259 -----AMDFEGLGDSNSNTLENSSDTVSIODSEEDMIYONSNEISIEQFTRREDOVE 1314
Db 2773 RGEKMEFIGIITDEBESQFPBOAETNKDIPBS-----VSTTGPIYSAALQTYKQ--- 2824
QY 1315 VLEPLKCELVGSESTGNCEDRLPVKGTBEANGK---RPSQOKULEEPRVVKCDOIKLN 1370
Db 2825 --EP-----STPNS-----KNEEAHTOLTUVHEPEOQOOLER---SRLSGSSSSS 2864
QY 1371 TTDKKNNENRESEK---KQGRSTFQIINGKDNKPKIYLKGECLKELISEKSVYVSGNVEPK 1466
Db 2865 HADREHRRERREKRRKRSOREQONQIHOKSSK-----VETK 2902

QY 1427 VNNINKIIPENDIKSLTVKESAIRPFIINGDVIMEDENERNSSETKSHLSSDAEGNYRD 1486
Db 2903 VDDNSGVMDDEAGRALAEQ-----LMSDEDTKRISB----- 2933
QY 1487 SLFTLPSTK---ESDSTOTTTPBASCPEBSNVQV-----EMELETSYVKKVTSSPIT 1537
Db 2934 --EATPSTATVYRSDMTDVFREFSDN--EDNNSVMTKQGVKSEQOEQHKSKDKKKRKRK 2990
QY 1538 SEESNLSNDFIDENGLPIKKNENVNGESKRTVITEVYTTMTSTVATESKTVIKVEGDK 1597
Db 2991 EERQEKLL-----QOQKRSLEPNVASTSAPPTPGKLYVNOVAASK 3031
QY 1598 QT-----VVSSTENCACKTV-----TTTTVYTKLSTPSTGSDIISVKEOS 1640
Db 3032 HADLOLDAKHISBPYCKPSPSLPCLIGDDDDDLAHTPKAKPTTPPSRGNDGLTPREKR 3091
QY 1641 KTV-----VTTTVTDSLT-----TTGTVLVS----- 1662
Db 3092 RLISPIKPTPIANSSTLSTQSAETPVSSGTVISSSALATPTSTYAGVSAAPGLDNP 3151
QY 1663 -----MTVSKYST-----RDKVKLKFGRPK 1685
Db 3152 TSASAOCKKKESTPGFDGOLDRIBESAVOSISAEFNSILLDNINADEKTIIVASPPRA 3211
QY 1686 TRSGTALPSYRKEPVTKSTK----- 1707
Db 3212 TKPLDKLEESKSVTTSQETESAVSALLGESFGTSTTDYSLDGMENSVNELETPL 3271
QY 1708 FVLPNDLKLAKLAK-----GIREVPYFNNAKPAIDIMVYBSP-----RPFPGIT 1753
Db 3272 VIAEPDEALAKAIETAGEPASILEEP---EMEBERAEPDEPPEAEIESEPEVEVL 3327
QY 1754 WRVRL--QTVKSLAGVSLMLRLMASLRMDMAKVPBGSGSTETSETTEITTELIRK 1812
Db 3328 DPBELKAVOSLKHEDM-----DIKADTPOSEBDLOIDI--DIFENDE---AD 3372
QY 1813 DVGPRGIRFEYCIRKIIICPIGVETPREPTPOKGLRSSALRPKRPETKQGVIIET 1872
Db 3373 SSGP-----SKIDETVQSSSPKESISNNS-----PREFETANIIDI-- 3409
QY 1873 WVAEELLEMEIARFARVKEKAQAVEQAKKRLBOKR-----TVLATSTTSPTSS 1926
Db 3410 -----PNVESQPKLSNESSTPOSVITKPLFELDPKTVPAGLRPPSEKIE 3453
QY 1927 TSTISPAQKVAVAPISGSV---TTGFKVLTTRKVGSPATVTRQONKNEFHQFATVWKQ 1982
Db 3454 PPTISKIQPLVQPVQVULPARHSTGSGISANSVINLDSNVSSCSNTSAASATASASA 3513
QY 1983 QNSGVVQVQKVLGILIPSSGTSGQOTFTSFQ--RATVVTIRPNTSGSGGTTSN---SOV 2038
Db 3514 SISGSPETASQNMAM---POASTPKOGFTTPOQAIPTOSLIMOPPTISIPROTGHFAVPOM 3570
QY 2039 ITGPOI-----RPG--KMTVITPLQOQSL---GKAIRRT-----PVWOPGAPQ 2078
Db 3571 VLSBPQHHPQOQGVYMGIAIRAPSPHSLPSRGGKAVQSLVQGLSPVGRWVQAP--SPQ 3629
QY 2079 VMTQIIRGQPVSTAVSAPNTVSSTPQCKSLTASITSNIOSSA---SOPRPOQGVOKLT 2135
Db 3630 ---QVQOTOQOQHALLITSPQSNISPLASPTTYRLVSSNSPTTKVANSYQORNO----- 3680
QY 2136 MAOITLTOGHGNGOGLTVVIOGOGOTTGLOLIRPOGVTVLPBPGOOLMOAMPNGTVOR 2195
Db 3681 -----VPOQSPKSAVAVUTTPQMLMT---PLQKMPFIQVPHNPT-- 3717
QY 2196 FLTPPLATTAATTASTTTTYYTTAAGTGBOROKSLSPQOMVHODKTLPRQOSSSVGRAPA 2255
Db 3718 --IISKVTVVPOQOATQOVAVSPPPLS-----LRPHKNVH-----LNAHQN-----QO 3759
QY 2256 OPOF--AOPSAKPOPOQOPAPPEVQTOPEVOTQTTVSSHVSEAPRTIAOSSKPOVA 2314
Db 3760 QPOVIAMTKAHQHQHMOQIMHOQMTORQOHMO-----QOQLHGOSO--QITS 3805

QY 895 WS-----WISKTHYRFPYKPLGNTNVYRKSLGCTKNNNDENDESDKRCSPKKIKIE 951
D 2475 DSDJTEEDHQH-----RNSGSSFFHGR--YPELSSDDDDDETHQRIKIS----- 2517
QY 952 PDESEKDEVKSDAAKAGADONE-----MDISKITEKKDQVKELLSDSDRCKEEMEYD 1006
D 2518 -----SDSDAEHFGODNQGASTLADANRVQRQO-QUNLRJCDDDDSD--SED--EIR 2563
QY 1007 DDMKTESHVANCQESSQOVVYVNNSEG-----FHLRTSY-----KKTTSKSKIDGL 1051
D 2564 RNVMKSHFEGRKNSNSTRIASDESQOPARDLTIKOEHPHAPAEIKREKROSDDEQOKER 2623
QY 1052 ERRIKQFTLEEKORLEKIKLEGIKIGKTSTNSKNLSESPVITAKEGCQSDSMROQ 1111
D 2624 SRHDSNSSIEER-----KLTEREI-----KTELGFYNSEFYITGGLKKEYSPETRKHHK 2675
QY 1112 SPNANDQPEDLJQGSQSDSS-----VLKMSDPS-----HTTNKLYPKDRVLDVSI-- 1159
D 2676 SKRR-----LKSSSTADTSAAQTPPLVMTPLTPSIFDVHSSSECKTKFDNFDLKTCEC 2727
QY 1160 -----RSPETKCPKONSIEENDIEEVSDLASGQEPYTSKTKGNDFIDDK 1206
D 2728 SSJPLEISAGERKHKREKREKREKLNMTATVPN-----SPTTNDT-----SSE 2773
QY 1207 LASADDIGTLICKNKKPLIOEESDTIYSSSKSALHSSVPKSTMDRATPLSR----- 1258
D 2774 KLSKEFHRLKSKSKSMNSCNTKIYNSSGA-H--PSTPSLPTATPTASASTAQTOK 2829
QY 1259 -----AMDEBKLCQDSESNSTLENSDPTVSIQDSSEDMIVONSISIQFTREDOVE 1314
D 2830 RGECKMEFIFGIISDEEESQFPEQAEFNKDIIPPS-----VSTGTGPVSAALQTYKO-- 2881
QY 1315 VLEBLKCELVSGSTGCEDRLPYVGTEANGK---KPSQKKLEERPVCXCDQJLKN 1370
D 2882 --EP-----STPNS-----KNEFAHJQLVHEPEQOOLER---SRLSGSSSSS 2921
QY 1371 TTDKNNENRESEK---KQORTSTQINGDKNPKIYLKGECLKEISERVVSGVNEPK 1426
D 2922 HADREHRRRREKREKREKQREQOQIHQSKK-----VETK 2959
QY 1427 VNNINKIIPENDIKSLTVKESAIRPFIINGDIVIMEDFERNSSFTKSHLLSSDAEGVYR 1486
D 2960 VDDONSVDMEAGRALEAQ-----LMSDEFTKPISE----- 2990
QY 1487 SLELTPSTK---ESDSTQITTPSASCESNSVNOV---EDMELETSEVKKVTSSPIT 1537
D 2991 --EATPPTATYRSDMTDVRFSDN-EDNNSVDMTKQGVKSEQEOQKSKDKKKKKRKS 3047
QY 1538 SEESNSLNDPIDENGILPINKENNVGESKRTVITEVTMTSTVATESKTVIKVEGDK 1597
D 3048 EEQEKELL-----QOORRESLPMVASTSSAPPTPGKLTIVWQVQASK 3088
QY 1598 QT-----VYSSFENCASKSV-----TTTTTTVKLSPTSGSGVDIISVYEOS 1640
D 3089 HADJQLDAKHHSPPVCKPSPSLPCLIGDDDDALHTPKAKPTTPSRGMDGLTPSRKPK 3148
QY 1641 KTV-----VTTTVDLSLT-----TTGGTLVYS----- 1662
D 3149 RLISPIRPTPIANSSTLSLQSAFTPVSSGTVISSALATPTTSTAAGVSAAPGLDNP 3208
QY 1663 -----WTYSKEVST-----KDKYKIMKFSRPK 1685
D 3209 TSASAGKCKKESFIPGFDQLDRISASAVOSISAENFNSLNDNIADERKIPVASSPRA 3268
QY 1686 TRSGTALPSYKRYTKSTKK----- 1707
D 3269 TKPJDKLEBSKSVTISOETESAVSALLGSEFTSTTDYSLDGMEMSSVNFLEPTL 3328
QY 1708 FVLNDDLKTLARK-----GGIREVPYFNNAKPALDIWYPS--RPTFGIT 1753
D 3329 VIAPDEEALAAKAIEIAGEPASIIEBP-----EMEPREAREPDPDEAEIESEPVVEVL 3384
QY 1754 WRYRL-QTVKSLAGVSLMLRLIMASLKMMDMAAVPPOGGSTRTJETSETTTTETIIRK 1812

D 3385 DEFEINKAVQSLKHBDM-----DIKADTPQSERDQLDIT-DTEENPDE---AD 3429
QY 1813 DVGPRGIREVYCIKRICIGVETPKETPRPQKGLASSALRRKRPRTPOGQVLIET 1872
D 3430 SSGP-----SLKIDETVQSSSPEKISNNNS-----PTEPANIDIT-- 3466
QY 1873 WVAEERLELMEIRAEVEREKEKAQAVEQAKRLEQOKP-----TYIATSTSPST 1926
D 3467 -----PNVESQPKLSNESTPQPSVITKLPFLDPRKTVAPGLPSPVKIE 3510
QY 1927 TSTISPAQKVMVAPISGV---TTGKMWLTTKVGSFAYVTFOONKFNHOTFAIWWQ 1982
D 3511 PPTISKLOPVLQVQVTVLPAHSTGSGISANSVINLDSNVISCSNSTASASATASASA 3570
QY 1983 QNSNGVVOVQOVULGAIPESTGTSQOTFTSPR-PTAYVTIPMNSGGSGGTTSN---SQV 2038
D 3571 SISFGSPNASQNAM---PQASPPKOGPTPQOALRQSLIMQPTIISIPQOTPHAVQOM 3627
QY 2039 ITGPOI---RPG--MTVIRTPLOOSTL---GKAIIRT-----PVWVQBARQ 2078
D 3628 VLSPOSHNPQOQGYVMVGIARSPHSPHSPBGRVAAQSLVQGLSPVGRMVSQD-SPO 3686
QY 2079 VMTQITRGVSTAVASAPRTVSGTQOKSLGATSTNIOSSA---SQPRPQOGCVKLT 2135
D 3687 ---QVOOTQOQHALLTSPQSSNISPLASPTTVLSSNSPTTSKVNSYQPRMQO----- 3737
QY 2136 MAQLPQLQHGNGGLGVVIGOGGOTTQOLIPQGVTVLPGPQOLMOAMRGTVOY 2195
D 3738 -----VPOQPSKSVABEYOTTQOLMI---PQOKTPIQVPHNR-- 3774
QY 2196 FLETPLATTATATSTTTTSTTAAGTBOQOSKLSPOMQODKTLPRPAOSSSVGRAPA 2255
D 3775 -ITSKVTVQPOQATQSOVASSPRLGS-----LPHKNVH-----LNAHQN-----QO 3816
QY 2256 QROT-AQBARQROQPOPSRQOREVQOTQOTTVSSHVPEAQRTNAOSSKROYAA 2314
D 3817 QROVIAKMTAHOHQHMOQRFMOHQTOHQHMO-----QOOLHGSO---QITS 3862
QY 2315 QSQPOSNQVQSGPVRVQVSPQTRIRPSTPSOLSPOQ-QSOVQOTTSTORPIQPHSLQTP 2373
D 3863 AROHQHQAQHQQAQOQOQHNNQOHLNQOHLQAQOHPRTQKHAQOQOFMQ-QIOQHOSQO-- 3918
QY 2374 SQGQPOSQPOVQVS---STQTLSSGQTLNOVSVSPBRQLOLQOQPR-QVIAVPLQLOQ- 2428
D 3919 -QHQVQOQOHAQOQOHLSDOQHOSQOQLOQO---QHQAQOQOQLOQLOKLOQMHGPOQOQKS 3973
QY 2429 -----VQVLSQISOVVAAQIOAQOQSGVPOQIKLQPLTQOQSSAVQTHQIONV 2476
D 3974 PQGVGHGSGSTIFASQOHNQOLPAR-----GVPOQ---QHPQOLSHSSPCKPRMTIVSV 4024
QY 2477 -VTVOAASVOEOLQORVQOHLRDOQKKQO---QIEINV--NTP 2513
D 4025 NQGVOPRALITRVGSHSOPRQOQOLPHQOSSGHPHQOQLSPGANLPLQTLPLMYONTP 4084
QY 2514 SKLLIVEIT-QOQV---VAKHNAVIEHLKOKKSMRAREENQMTVCNOVMKYTLDKI 2569
D 4085 -KIIVOOHIVAQOVPRPQOTQGNALHYPONQGDSTPRGH----- 4123
QY 2570 DKEEKOAKKRRKRESEVQOKRSKOMATKLSAL-----LFKHKEOLAELIKKRALLDK 2622
D 4124 -VEPTPAMSNQKSESVSVIRPTPTTGLAVISANTVGSILTEENLIKISQPKQDELBIQ 4182
QY 2623 DLQIEWO-----ELKRDIKIK-----EKDLMQLOATAVAACPRVTYVLAAPRP 2670
D 4183 DSK-EVDSQVSAKEAVINQSVIKKLDTPLASKAKAKAVEQOAL-----AP 4226
QY 2671 PPSPPRPQVQNGHGLSTPLPVAOSQKRRKEEKSSSKKKKM 2715
D 4227 APIPNQPGNQ--SMAQETALPTTSMVNSNNDHDEDETETRQL 4269

RESULT 13

09NHNT
ID 09NHNT PRELIMINARY; PRT: 5554 AA.
AC 09NHNT;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Split ends long Isoform.
SPEN OR CG18497.
GN
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RA Kuang B., Wu S., Shu Y.-A., Luo L., Kolodziej P.;
RT "Split ends encodes large nuclear proteins that regulate neuronal cell
RT fate and axon extension in the Drosophila embryo."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF221715; AAF34661.1; -
DR HSSP; P09651; 1HA1.
DR FLYBase; FBgn0016977; spen.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.
SQ SEQUENCE 5554 AA; 59918 MW; 4037E27833D0C622 CRC64;

Query Match 4.1%; Score 613.5; DB:5; Length 5554;
Best Local Similarity 19.5%; Pred. No. 5.2e-16;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;

QY 305 DCAVEIQKPKFYIHPEYIGYDRSRKRYWFLNR-----RLIIEEDT 344
DB 1906 DTAQKAENQ---RHEKKKKRQERREKDLKQYVEREKKDKAQQEEREKKKEEK 1962
QY 345 ENEWEKTKWYSTVQALAEIDLDKDYWEALCKILEMSEBEIHRMDTE-DLTNKR 403
DB 1963 EREKEREKQEDREK-----KEREERELRE--KEQRDKQEKKEIREKDLREKQ 2009
QY 404 GSNSFLAANBEILESTIRAKKGDIDNVKSPETPEKDKENFENDSKDKNREEDSL 463
DB 2010 -----REKDNREKREKDLKDLREKREKQERKELHREKQERREHR 2051
QY 464 EKSDSDKTDDDEP--QKSESEPT-----YDKGNSVANLGDWTATSEETSPS 513
DB 2052 EKESRRAMDVEQSGRGGRMELSYOKSKMDIAGEASLALDQNHKKAAMDITAGT 2111
QY 514 EGRSPVGLSTPDSSNAKERVASLPRQVPEEPN--TCESSTSA-----TTTSI 564
DB 2112 PGASP-----STP--SDTTPKERSRKLSRNSPVLRLHRRLLSQEENHSAAGGCGSSH 2164
QY 565 QPNLENSSSSELSSQSEKAKADDPENGERS--HTPVSTQETVDDTFSEKTEGSL 622
DB 2165 QIHNEDEVKIRIMNSQNIIVSHSSNQRLNDRDSKEHKSSEFKED-----KNSSHIS 2217
QY 623 ESPAGKAGSGSTRITLRLRMPDSKLSQKQOYAAAHNAKMLEKEGEVLYVNSQEI 682
DB 2218 RPHCGGSGSASSSKHNH-----RDKHKGSSASSETNNSI---EYVVDISQTK 2265
QY 683 SRLSTKKEVINKGINNYFKLGQBGKYRVYHNOYSTNSPALNK-----HQRH- 729
DB 2266 HNLMTSEELQSHQ-----PKREKEREHFSHANSSSSRHKSRDHHRREKKRHSVA 2318
QY 730 -----EDHKRRRLAHKFCLT-----PAGEFKWNGSVGKVLITSTRLITTOLENNIP 779
DB 2319 ESTNTDEHTEQOHNPRIISAAGSGSAGELSSAATNTSGKLLHQHRRRSVERKSSGS 2378
QY 780 SSFHPNASHARAWMIKAVQCSKPREFALALALECAVKVYVMLPIRREFLGHTLRHM 839
DB 2379 DEGHSSSKSLRAKALIMLMSADSDTDDA-----SKKSI 2413
QY 840 TSIERE-----EKEKVKKKKKEKEEETMOATWVTKYTPVKHQVKKQKGEEYRVTYGG 894

DB 2414 FDIPDDCPNVSMTDKVKARSKNMQROAEKKIKAKFS-----OLKQSRAKKRSTSYDG 2468
QY 895 WS-----WISKTHYRRVPLPGNTIVNYRKRSLLEGTKNNNDENNDEEDKRCSPKIKIE 951
DB 2469 DSDTEFEDRQH-----RNSGSSSFHGR--YPLGSSDDDDDEETHORIS----- 2511
QY 952 PDSKEDEVKGSDAKGAQONE-----MDISKITEKKDQVVELDSDSDKPKCEPMEDV 1006
DB 2512 -----SDSAEHLGGQDNGGASGLADANRYRQMQ-QNLRRLCGGDS---SED--EIR 2557
QY 1007 DDMKTESHVNCQESSQVDVYVNSG-----FHLRTY-----KKTGKSKLDGILL 1051
DB 2558 RVMKSHSFGRKNSNSTRIASDESQSPAPDLTIKQHPRIAPAEIRREQLSDEQKFK 2617
QY 1052 ERRIQFLTEQRKREKIKLEGIGIKGTSTNSKNLSESPVITKAKEGQSDGSMRQO 1111
DB 2618 SRHDSNSSIEER---KKTEREI---KTELGFDFNNSSETTYTGKLEYSPEPKKHK 2669
QY 1112 SPNANNDOPEDLIQGCSQSDS-----VLRMSDPS---HTTNKLYPKDRYLDVSI-- 1159
DB 2670 SKRR-----LKSSSPADISAAQTPLYMTPLPFSIFDVHSSSECKTKPFQNFDLKTEC 2721
QY 1160 -----KSPETCPKQNSIENDIEEKVSDLASRQGPYKSKTKGNDFTIDSK 1206
DB 2722 SSIPLEISAGERKHKHKKKKREKRLRMTEATVFN-----SPTTNDT-----SSE 2767
QY 1207 LASADDITGLCKNNKPLIQQESDPIYSSSKSALHSSVPKSTNDADAPLSR----- 1258
DB 2768 KLSKERHRLKSKSKSKMDNSCNTRYINSSGA--H--PSTSPSLPAPTPASPTAQTGSK 2823
QY 1259 ---AMDEGKLGCDSESNSTLENSSDPVSJODSSSEEDMIYQNSNESTSEQFRTQDVE 1314
DB 2824 RGEDMEFIPIGLISDEESQPEQAETKNKIDIPSS-----VSTTPRIYSALQTKQ--- 2875
QY 1315 VLEPLKCELVGSESTGNCEDRLPVKGTGANCK---KPSQOKLEERPVNKCSDQIKLN 1370
DB 2876 --EP-----STPNSS-----KNEBAHIQTLVHEPEQOOLER---SRLSGSSSSS 2915
QY 1371 TTGDKNNMRESEK-----KQORTSFOJNGDKNKKIYLGKCLKEISESRVSGNVPEK 1426
DB 2916 HADREHRRERKREKRRKRSQEQOQNIHQSSK-----VETK 2953
QY 1427 VNINIKIIPENDIKSLTVKESAIRPFIINGDVIMEDFENRSEETKSHLLSSDAGNRD 1486
DB 2954 VDDDSVMDDEGRALQ---LMSFDKRPISF----- 2984
QY 1487 SLETLPSRK---ESDSITQTPPSACPSNSVNOY-----EDMEIETSEVKKYTSPT 1537
DB 2985 --EATPSTAAATYRSDMTDVPFRSDN--EDNNSVDMTKQGVKSEQOQHRSKDKKKKRSK 3041
QY 1538 SEESNLENDPFDENGLEINKNNENNGESKRTVITTEVTMTSTYATSKYIVIKYKEDK 1597
DB 3042 EEQKELL-----QOQRESLPPNASTSSAPPTGKLTIVNOQASK 3082
QY 1598 QT-----VVSSTENCAKSTY-----TTTTVTVKLSPTSGSGVDIISVKQOS 1640
DB 3083 HADLODAKHHISSPVCKPSPSLPCLIGDDDDDLHTKAKAPTPPSSNGNGILTPSRKRP 3142
QY 1641 KTV-----VTITVTDLSL-----TTGGTLVYS----- 1662
DB 3143 RLSPIPKPTPIANSSTLSTQSAETPVSSGTVISSALATPTTSSTAAGVSAAPGLDNSP 3202
QY 1663 -----MTVSEYET-----RKVYLMKFSRKK 1685
DB 3203 TSASAQCKKESFIPGFGOLDRISESAYOSISAEFNSTLNDINADPKIPIVASPPRA 3262
QY 1686 TRSGTALPSYRKFEVTKSTKK-----SI 1707
DB 3263 TKPLDKLEESKRYTISQEEIESAVALLGSEFGISSTTDVSLGMDKSSVNELETPTL 3322
QY 1708 FVLPNDDLKLARK-----GGIREVPFNYNNAKRALDIMPVSP-----RPTFGIT 1753

Db 1912 DPAKAEKNO---RHEKEREKREKDLRKOVEREEDKRAQOEEREKEDKAKAEK 1968
Qy 345 EBNENEKITYSTKVOLAELIDCLDXYDAELCYLLEKEEREIRHMDITE-DLTNKAR 403
Db 1969 EBEREKKAODEBERK-----KEREERELRE--KEQORKEKEREKELREKREKEQ 2015
Qy 404 GSKSKSLAANBELLESTRAKKGDIDNVASPEFTEKDKNTENDSDAKENREEFEDSL 463
Db 2016 -----HERDNREKELDKDLREKEMREKEQEREKELHREKQOREHER 2057
Qy 464 EKDSODKTPDDPE--QGKSEEPTE-----VGDKNSVANLGDNTTNNATSEETSPS 513
Db 2058 EKEQSRADVDQEGGKMBRELSTYOKSKMDLAGASSITLIDCQHNKENAMDTAQT 2117
Qy 514 EGRSPVGLSETPDSSNMAEKVASELPDQVPEPNK--TCBSSNTSA-----TTTSl 564
Db 2118 PCASP-----STP--SDNPKERSKRLSRNSPYRLHKRLRLSOGESNHSAGGGSCGSSH 2170
Qy 565 QPRLSNSSSSELNSQSASAKAADPENGERS--HTVYSIOEYVGDTSEKSTGELS 622
Db 2171 QJHHEBYKRIKRENSQNTSVHSSNORLDRDSEKSSFEKED-----KNSSSHIS 2223
Qy 623 ESPGACKGASGSTRITRIIRNPDSKLSQLKSOQAAAHBANKLFKEGEVLVNVQGEI 682
Db 2224 RHGCGGSSASSKHNHR-----RDKHNOGASASITETNST--EVVVDPISTQK 2271
Qy 683 SRLSTKEVIMKGINNYEKLQEGKYRYVHNOYSTNSPALNK-----HOHR----- 729
Db 2272 HNLNTSEELQSHQ-----PKREKEREHFSHANSSSRHKSXRDHNHRRKKHNSVA 2324
Qy 730 -----EDDKRRHLAKCLT-----PAGEFKNGSVHSGKULTITLRLTTOLENNIP 779
Db 2325 ESTNNTDEEHTPOOHNHRRIRISAAGSGASELSAATNTSGKLHNOHNRSVYERKSSRG 2384
Qy 780 SSFPHNMAASHRANWIKAVOMCSKPREPALALALECAKPVVLMIRFELGTBLHRM 839
Db 2385 DEGHSSSSSLRAKMLMLSSADSDTDDA-----SKKHST 2419
Qy 840 TSIERE-----EKEVYKREKQOEETEQOATVKTPEPVKQHVOKQEGEYRVGYG 894
Db 2420 FDIPODCPVNMVDYKAKSCNMQOABEKKIKAFKS-----QLKQSRKKKRRSYDG 2474
Qy 895 WS---WISTHYRVPKLPQNTNWKSLBSTKNNMDENDESKRCSRPKKIKIE 951
Db 2475 DSDTEFEDQH-----RNSGSSSFHGR--YFGLSSDDDDDEETQORRIS----- 2517
Qy 952 PSEKREXVAGSAKGAQONE-----MDISITEKKDQVYKELSDSDKPCKEPMEVD 1006
Db 2518 -----SDSDHEHGQDQCASTLADANRVQOQO-QNLRRLCDGDD--SED--EIR 2563
Qy 1007 DDMKTESHVNCQSSQVDVNVNSEG-----FHLRTSY-----KKTKSSKLDGLT 1051
Db 2564 RVNMKSHHGKGRNNSNTRIASDSESGOPARDLTIKQEHPIAQAQELKREQSLDEQOKR 2623
Qy 1052 ERRIKOTFLEKQORLEKIKLEGIGIKGTSTNNSKNLSBPVITAKAGCQSDSKRQEO 1111
Db 2624 SMDSSNLSIEER---KLTEREI---KTELGDFFNSSEYTYGKLEKESPEFRKKHK 2675
Qy 1112 SPNANNDQREDLLOGSOSDSS-----VLMSDPS-----HTNKLKXPDRVLVDVST 1159
Db 2676 SKRR-----LKSSTADTSAOQPLVMTPLTPIFVHSSSEKKTFFDNFDLKTETC 2727
Qy 1160 -----RSPETKCPKONSIEENDIEEKVSLASRGQEPSTKRGKGNDFIDSK 1206
Db 2728 SSIPLEISAGERKHKERKEKKREKRLNMTFATVPN-----SPTINDT-----SSE 2773
Qy 1207 LASADDIGLICKNNKPLIOESDITVSSSKSLHSSVPKSTINDRATPLSR----- 1258
Db 2774 KLSKEERHRLKRSKSKSDMNSCNTKIYNSGA-H---PSTSPLPATPTSAESTAQTOK 2829
Qy 1259 ---AMDFGKLGQCSSESNSTLENSDYVSIQDSSSEDMIVONSNSISISQFETRRQDVE 1314
Db 2830 RGEDEKMEIFGLISDEESQFPEOAEITKNKDIIPSS-----VSTTGPVSAIQTAYKO--- 2881

Qy 1315 VLBPCLCELVSGSTGNCEDRLFYKGTENGK-----KPSQOKLIERPVPVKCSDQIKLN 1370
Db 2882 --EP-----STPNS-----KNEEAHIQTLVHEPEOQOOLR---SRLSGSSSS 2921
Qy 1371 TTDKNNENRESEK-----GQORTSTQINGKNNPKRYILKGECLKELSESRVNSVNEPK 1426
Db 2922 HADREHRRREKREKREKRSQREOQNOIHOKSK-----VETK 2959
Qy 1427 VNNINKIIPENDIKSLVKESAIRPEINGDIMEDENERNSSTKSHLLSSDAEGNYRD 1486
Db 2960 VDDNDVNDMEARALEAQ-----LMSDDTKRIS----- 2990
Qy 1487 SLELPSTK---ESDSTQTTTPASCPESNSVNOV-----EDMELETSEVKKVTSSPT 1537
Db 2991 --EATPSTATYRSDMTVDYFRFSDN--EDNNSVDMTKQGVKSEQOQHKKSKDKKKKRSK 3047
Qy 1538 SEESNLSNDFIDENGLPIKNNVNGESKRYVITEYMTTSTVATETSTYKVEKDK 1597
Db 3048 EERQEKLL-----QOQRRESLPVAVSTSSAPPKGLTVNVOAASK 3088
Qy 1598 QT-----VVSSTENCAKSTV-----TTTTTVTKLSTPSTGSGVDIISVKEOS 1640
Db 3089 HADLODAKHISSPPCKPSPSLPCLIGDDDDALHTPRKAKPTTPPSRGNDGLTPREKR 3148
Qy 1641 KTV---VTTTVDLSLT-----TTGGTLVTS----- 1662
Db 3149 RLSPKPTPTIANSSTLSTQSAETPVSGTVISSALATTPTSTRAAGVSAAPGLDNP 3208
Qy 1663 -----MYVSKEST-----RDKVKLKKFSRPK 1685
Db 3209 TSASAOCKKKESEFIPFGDGLDRISBAYOSISAEFNSTLNDIADDEPKIYVASSPRA 3268
Qy 1686 TRSGTALPSYKRFVTKSTK-----SI 1707
Db 3269 TKPLDLKEBSKSVTTISOETESAVNALGESFGTSTTQYSLDGMDEMSVNELEPTL 3328
Qy 1708 FYVLDLKLKARK-----GGIREVPYFNANPAIDIMYPSP-----RPTFGIT 1753
Db 3329 VIAPEBEALAKAJETAGEPASILEEP-----EMEBERAEEDPOPEABEISEPVAYEL 3384
Qy 1754 WRYRL-QYVSLAGVSLMLRLMASLRWMDMAKVPPGSGSTTESEYETITTEIIRK 1812
Db 3385 DPBELKAVQSLHEDMM-----DIKADTFQSEKDLOIDT--DREBNDE--AD 3429
Qy 1813 DWGPGYRFEYCIKLIICPIGPEPEKPEPTPOKGLRSSALPKRPREKOTGPVIEET 1872
Db 3430 SSGP-----SIKIDETVOSSSPKSTISNNS-----PYPRETANIDI-- 3466
Qy 1873 WVAEELELMEIRAFARVEKEKAKOAVEQAKKRLQOKP-----TVIATSTTSPTSS 1926
Db 3467 -----PNVESQPKLSNESPQPSVITKLPLFDLPKTVAPGALPSPVKIE 3510
Qy 1927 TSTISPAQVMAVAPISGV---TTGTKVLLTKVSPATVTVQOKNHNQGTATVWQK 1982
Db 3511 RPTISKLOOPVLOPVOTVLPAPHSTGSGISANSVINLDSNVISCSNTPSASATASASA 3570
Qy 1983 QNSGVVVOQOKVGLGIPSTGTSQOTFSTFOP--RATVIVIRPTGSGSGTSTN---SOV 2038
Db 3571 SISFGSPASQNAM---POASTPKQCPITPOQAIQOSLIMOPTTISIPQTHFAVPOH 3627
Qy 2039 ITGPOI---RPG--MTVITPLQOSTL---GKAIIRT-----PVWVQPGAPQ 2078
Db 3628 VLSRQSHNPOQPTGVVUGIARPSHPLSGGAVQSRVLVQGLSPVGRMVQSP--SPQ 3686
Qy 2079 VMTQIIRGQPVSTAIVASAPNTVSTPQOKSLTSATSTSNIOSSA---SOPRPOGQVKLT 2135
Db 3687 ---QVOTOQOAHILITSPSSNISPLASPPTTIRVLSSNSPTTKVNSYORRNO----- 3737
Qy 2136 MAQLQTLQGHGNOGLITVYIOGOGOTQOOLIPQGVVLPBPGQOLMAAMPNGTVQR 2195
Db 3738 -----VPOQSPKSVAVAEVOTTPOLMTI---PLOKMPPIOVPHHPT-- 3774

```

QY 2196 FLEPTPLATTATTTTSTTAAGTGROKSLSPOMOHODKTLTPRAOSSVGPAXK 2255
Db 3775 -ITSKVVTVQPOQATOSQVASSPPLS-----LPPKKNVH-----LNHQN-----QY 3816
QY 2256 QPQT-AQPSARPQQTQOPSPAPQEVOTQEVOTQTTVSSHVSEAPTHAOSKPOVA 2314
Db 3817 QPQVAKMTAHNQHQHMOQFMHQMIORQOHMQ-----QQQLHGQSOQ-QITS 3862
QY 2315 QSQPOSNVQGSFVRYVQSPQTKIRPSTPQSLPQ-QSQVQTTTTSQPIPIQHTSIQIP 2373
Db 3863 APQHONHQHQAOQOQOQHNNQOHLNQOHLAQOHPQKQHQAOQOQFNO-QIQOHSQOQ- 3918
QY 2374 SQGQPOSPQVQVS---STQTLSSGQTLNQVSVSPSRPOLQIQOPQ-QVIAVQPOLQOQ- 2428
Db 3919 -QHMQVQOQNAQOQOQHLSQOQHOSQOQNLQ---QHQAQOQOQLOIQIKLOMHGPRQOQKS 3973
QY 2429 -----VOVLSQIQSVQVVAIQAOQGSVPQOIKQLPIQIQOSSAVQTHQIQNY 2476
Db 3974 PQGVHLAGGSTSIFASQOHSQILPAP-----GVPOQ---QHPOQLSHSSPCKPNTLVSV 4024
QY 2477 -VTVQASVQEBQIQVQQLRDQOQKKKQO-----QIEINV---NTP 2513
Db 4025 NQGVQPPALITRVGSHSQPNQOQQLPHQOSSGSHPHQKOLSPGQNLPLQTPLVQIQTNP 4084
QY 2514 SKLLIVEII-QKQV---VAKHNAVIEHLKQKSKMTPAREEENQRMIVCNQVAKYLLDKI 2569
Db 4085 -KIIVQOHIVAQNVPPRQPGQNAIHYPQNGQDSTIRPGH-----4123
QY 2570 DKEKQAKKRRKESEVEQKRSKQNAFKLSAL-----LFKHEQLREILKRLALLDK 2622
Db 4124 -VEPTPAMSAQKTSSESVIRPTPTTGLAVISANTVGSILTEENILIKITISQPKODELIEQ 4182
QY 2623 DLQIEVQ-----EELKRDKIKK-----EKLMLQAOATAVAACPPVPLPAPAP 2670
Db 4183 DSK-EVDSQVSKKEVNDISVYIKKLDTPLASAKAKRAVEQAI-----AP 4226
QY 2671 PPSPPPPPVQHTGLSTPLPVASQKRRKEEKSSSKSKKKM 2715
Db 4227 APIPNQPGNQ--SMAQETALPTSMVSNNSNHDTEDETFETROL 4269

RESULT 15
Q906C3 PRELIMINARY; PRT; 5533 AA.
AC 0906C3;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, last annotation update)
DE SPEN RNP motif protein long isoform.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyrotidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISO1.
RA Welleite E.L., Harding K.W., Mace K.A., Ronschaugen M.R., Wang F.Y.,
RA McGinnis W.;
RT 'Spn encodes an RNP motif protein that interacts with Hox pathways to
RT repress the development of head sclerites in the Drosophila trunk.';
RL Development 0-0-0(2000).
DR EMBL: AF188205; AAF13218.1; -.
DR HSSP: P09651; 1HA1.
DR FLYbase: FBgn0016977; spen.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rnm. 3.
DR SMART: SM00360; RRM. 3.
DR PROSITE: PSS0102; RRM. 3.
SQ SEQUENCE 5533 AA; 597114 MW; AFG60606DDEF269 CRC64;

```

Query Match 4.1%; Score 608.5; DB 5; length 5533;

```

Best Local Similarity 19.4%; Pred. No. 8,1e-16;
Matches 543; Conservative 406; Mismatches 1015; Indels 841; Gaps 117;

QY 305 DCVAELQKKNKPYIRNHEPIGDKSRKRYWFLNR-----RLIEEPT 344
Db 1912 DTADRAEKQNO---RHEKEREKEREKEDLRKQVREEREDKRAQOEEREDKRAKEE 1968
QY 345 ENENEKIKWYSYTKVQLAELIDCLDQWYEAELCKILEEMREIRNHMDITE-DLTNKA 403
Db 1969 EREKEREKAOEDREK-----KERERELE---KEQKQEQKEKELREKDLAEKQ 2015
QY 404 GSNKSEFLAANEELIESIRAKKGDIDNVKSPEETEKMETENDSKDAKNEEFQDGL 463
Db 2016 -----REDREREKELRQDKREKEMERKEQREKELHREKQDQREHER 2057
QY 464 EKSDDKTPDDPE--QKSEEPTE-----VGKQNSVSNLQDNTNATSEETSPS 513
Db 2058 EKQSRRAMVDEQEGRGRRRELISYQSKMDIAGASLITALDCQHNENAMDITAGT 2117
QY 514 EGRSPVGCLEETPDSSNMAKKVASELPQDVPREPK-----TCESSNTSA-----TTTST 564
Db 2118 PGASP-----STP--SDNTPKERSKLSRNSPVRLIKRLRLSQQSHNSAGCGGSGSH 2170
QY 565 QPNLENSNSSSELNQSQSEKAAADDPENGERS--HTPVSIQEEIVQFTSEKSTGEIS 622
Db 2171 QIHEDYVKRIRMENQNISVHSSNORLNDNRDSEKHSSEFKED-----KNSSSHIS 2223
QY 623 ESPGAKGASGSTRITRLRNPSKLSQLKQOYAAAHANKLFEKQEVLYVNSQGEI 682
Db 2224 RPHCGGSSASSSKHNHR-----RDKHHQGSASLITNSIT---EVVADPISQTK 2271
QY 683 SRLSTKEVIMKNNINNYEFLQGEQKYRVYHNYSTNSFALNK-----HQHR----- 729
Db 2272 HNLTSEELQSHO-----PKREKEREFSSHANSSSRHKSCKRDHNNHKKRHNSVA 2324
QY 730 -----EDHDKRRHLAKFCLT---PAGEFKNGSVHSGKVLITSLRLITTOLENNIP 779
Db 2325 ESTNTDEEHTPOQHNPRRIRISAAGSAGELSAATNTSGKLHNOHNRSVYERKSSRS 2384
QY 780 SFFHFNMAHNRANWIKAYQMCSPREFALALILDCAVPVYMLRIWREFLGHTRLHM 839
Db 2385 DEGHSSSKSLRAKLIMLSASDDDTDDA-----SKKHST 2419
QY 840 TSIERE-----EKEKYKKEKKEOEETMOATVYATPEVKHQVKKQGEYRVYGYG 894
Db 2420 FDLPDQPNVSMYDKYKARCKNMQQAEEKKIKAFS-----QLQSRAKKRRSTSYG 2474
QY 895 WS---WISKTHYRVYKPLRGNTNVYRKSLBETKNNMDENMDESOKKCSNPKIKIE 951
Db 2475 DSDTEFEDRQH-----RNSGSSSFHGR--YPLSSSDDDDDDETRQRRIS----- 2517
QY 952 PDESEKREVKSDAKKAGDQNE-----MDISKITEKKDQVKELDSDSOXPKCEEPMEVD 1006
Db 2518 -----SDSDAHEGQDQOQASTADANKVQMO-QNLRRLCDGDS---SED--ELR 2563
QY 1007 DDKTESHVNCQESQVDVYVNSEG-----FHLRTSY-----KKTKSSKLDGLL 1051
Db 2564 RNYMKSHHEKRNKSNSTRIASDSESQAPARDLITQENHIAQAEIKRQGLDSEQKFR 2623
QY 1052 ERRIKQFTLEKORLEKIKLEGIKIGKTSNNKSLSESPYITAKKECQSDSKRODO 1111
Db 2624 SRHDSNSTIER---KLTREI---KTELDFYNSSEYTYGKIKEPSPEPTRKKNRK 2675
QY 1112 SPANANDQPEDLLOGCSQSSS-----VLKMDPS---HTNKLKLYPRDVLVDSTI-- 1159
Db 2676 SKRR-----LKSSSTATSAQGPLVWTPPLTPISIPVHSSSEKTKTFDNEFDLTKEC 2727
QY 1160 -----RSPETCKPKONSIENTDEEVSYSLASGQPTKSKTKGNDFTIDSK 1206
Db 2728 SSTPLEISAGERKKHKEKREKELRNMTAEIVPN-----SPTINDT---SSE 2773
QY 1207 LASADDIGTLICNKKPLIOESDITVSSSKSALHSSVPKSTINDRATPLSR----- 1258

```


THIS PAGE BLANK (USPTO)

SQ SEQUENCE 810 AA: 91799 MW: 77E2C992FE5BE96D CRC64:

Query Match 27.3% Score 4082: DB 1: Length 810:
Best Local Similarity 98.7%: Pred. No. 6.1e-139:
Matches 777: Conservative 4: Mismatches 2: Indels 4: Gaps 1:

```

QY 1 MYSEEEEDGAEETODEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFSRSHST 60
DB 1 MYSEEEEDGAEETODEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFSRSHST 60
QY 61 YSTPGRKRRPRVHRPSPPLLEKDIPLPEPKSSSEDLAMPNEHINVAIAYEVLNFGTV 120
DB 61 YSTPGRKRRPRVHRPSPPLLEKDIPLPEPKSSSEDLAMPNEHINVAIAYEVLNFGTV 120
QY 121 LRISPRFDFCAALYSQSCITLMAEMHVVLLKAVLRERDSTSTGPGADLDKSVSTLY 180
DB 121 LRISPRFDFCAALYSQSCITLMAEMHVVLLKAVLRERDSTSTGPGADLDKSVSTLY 180
QY 181 FTIDGMPVFLRYVYCSDEKFEYHNVLPYQEAEDYPYGPVENKIKVLOFLVDQFLTNIARE 240
DB 181 FTIDGMPVFLRYVYCSDEKFEYHNVLPYQEAEDYPYGPVENKIKVLOFLVDQFLTNIARE 240
QY 241 ELMSEVIOYDHCRCVCHLGLDLCCETCSAYIHECVKPRLEVPDEMOCEVCAHKV 300
DB 241 ELMSEVIOYDHCRCVCHLGLDLCCETCSAYIHECVKPRLEVPDEMOCEVCAHKV 300
QY 301 PGVTDCAVEIOKNKPRIHREPIGYDRSRKRYMFLNRLLIEEDTEENKKIKYVSTKVO 360
DB 301 PGVTDCAVEIOKNKPRIHREPIGYDRSRKRYMFLNRLLIEEDTEENKKIKYVSTKVO 360
QY 361 LAELIDCLDKYWEALCKILEEMREIHRHMDITEDLTNKAAGSNKSLAANEILLES 420
DB 361 LAELIDCLDKYWEALCKILEEMREIHRHMDITEDLTNKAAGSNKSLAANEILLES 420
QY 421 IRAKKDIDNVASPEETENDKNTENDSKDAEKNEEFEDQSLERKSDDKTPDDPEQCK 480
DB 421 IRAKKDIDNVASPEETENDKNTENDSKDAEKNEEFEDQSLERKSDDKTPDDPEQCK 480
QY 481 SEEPTEVGDKGNSVANTLNTNTATSEETSPSEGRSPVGCLETPDSSNMAEKVASEL 540
DB 481 SEEPTEVGDKGNSVANTLNTNTATSEETSPSEGRSPVGCLETPDSSNMAEKVASEL 540
QY 541 PODVPEPKTCESSTNTATTTISIOPLNENSSSELSNQSSESAKAADDPENGERESH 600
DB 541 PODVPEPKTCESSTNTATTTISIOPLNENSSSELSNQSSESAKAADDPENGERESH 600
QY 601 PVSIOEIVGDTSEKSTGEISPGAGKAGSGSTRITRLRNPDSKLSQVAAAA 660
DB 601 PVSIOEIVGDTSEKSTGEISPGAGKAGSGSTRITRLRNPDSKLSQVAAAA 660
QY 661 HEANKLFEGKEVLYVNSQGEISRLSTKKEVIMKGINNYFKLGQKRYVHYNOYSTNS 720
DB 661 HEANKLFEGKEVLYVNSQGEISRLSTKKEVIMKGINNYFKLGQKRYVHYNOYSTNS 720
QY 721 PALNKHQHRHDHDKRRHLAHKFCLPAGEFKWNGSVHGSKVLITSLRLITOLENNNIS 780
DB 721 PALNKHQHRHDHDKRRHLAHKFCLPAGEFKWNGSVHGSKVLITSLRLITOLENNNIS 780
QY 781 SFPHPMW 787
DB 777 TSLHPSF 783

```

RESULT 2

TRX_DROME STANDARD: PRT: 3726 AA.
ID TRX_DROME Q27255, Q27327;
AC P20659; Q27255, Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila melanogaster (Fruit fly).

```

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
RT in Drosophila, encodes a protein with zinc-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
RN [2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=792496;
RA Sedkov Y., Tiliib S., Mizrokhi L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
RT Drosophila embryogenesis than is the Antennapedia complex, correlating
RT with a bithorax-like expression pattern of distinct early trithorax
RT transcripts.";
RL Development 120:1907-1917(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RT between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
RT directly regulates the region-specific homeotic gene fork head.";
RL Genes Dev. 8:2478-2490(1994).
CC -I- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -I- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
CC ZINC.
CC -I- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -I- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -I- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31617; AAA29025.1; -
CC EMBL: 250152; CAA90514.1; -
CC EMBL: 250152; CAA90513.1; -
CC EMBL: 231725; CAA83516.1; -
CC EMBL: 231725; CAA83515.1; -
CC PIR: A35085; A35085.
CC HSSP: P20393; 1A6Y.
CC TRANSFAC: T00850; -.
CC Flybase: FBgn0003862; trx.
CC InterPro: IPR003889; Flyrich.C.
CC InterPro: IPR003888; Flyrich.N.
CC InterPro: IPR003616; PostSET.
CC InterPro: IPR001214; SET.
CC InterPro: IPR001965; ZnF_PHD.
CC InterPro: IPR001841; ZnF_ring.
CC Pfam: PF00628; PHD; 3.
CC Pfam: PF00856; SET; 1.
CC SMART: SM00542; FYRC; 1.

```

DR SMART: SM00541: EYRN. 1.
DR SMART: SM00249: PHD. 4.
DR SMART: SM00508: POSTSET. 1.
DR SMART: SM00184: RING. 3.
DR SMART: SM00317: SET. 1.
DR PROSITE: PS0280: SET. 1.
DR PROSITE: PS01359: ZF-PHD_1. 4.
DR PROSITE: PS0016: ZF-PHD_2. 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3599 3708 SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-SER.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VAAPSLIC 1 368 MISSING (IN SHORT ISOFORM).
FT CONFLICT 2025 2025 P -> PMLTSPKLGSLTHGGLLMLLLGVVRLKQGG
FT CONFLICT 2341 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 2392 G -> S (IN REF. 1).
SQ SEQUENCE 3726 AA; 400575 MW; D275650763DICF5 CRC64;
Query Match 3.2%; Score 476; DB 1; Length 3726;
Best Local Similarity 18.8%; Pred. No. 1.1e-09;
Matches 628; Conservative 419; Mismatches 1138; Indels 1154; Gaps 155;
QY 5 EEEEDGAEETQSEED-EEDEMEEDDDSDYPREMEDDDDASYCTESSFRSHST 60
DB 538 DEDDEGVTFRRNDSPEDONNAEDDDDDDEAEEDDDONEDNDEASAKSAETKS 597
QY 61 YSS-TPGR--RKRPRVHRPSLLEEDIPLEPKKSEEDLMPN 101
DB 598 AGADRDEDEKOLVDSHFVLPKRSTRSSRIITPKPKRLLEGAI-----STKKPLSLGD 651
QY 102 EHMNVIAI-----YELRNFGT-----VLRSPFRPE 129
DB 652 SKGRVFGTSSSSAGSTASTSTNLKKEFPNFGTLKPNSSAAGFVLRQRLQFQ 711
QY 130 DFCALVSOECTLMAENHVVLLKAVLREEDTSNTTGPADIKDSVNSTLYFIDGMTWE 189
DB 712 -----ADNOQATEFAPRACPTSPSAIKPANSLATSSFGSLATNSSTV-----TPTPS 760
QY 190 VLRYVCE--SDKEVHNHLPQEAEDRYGPEYENKI--KVLDFVLDOFLITINAEELMS 244
DB 761 ACST-CSAVVSKEY-----TOARKYGVACDVCKRFPKSKMTKKSISANSTAN-TS 810
QY 245 EGVLOYDHCHEVCHKGLDLC-----CETCSAVYHLECVAPLLEVPEDMEQCEVC-- 295
DB 811 SSGSQY-----LQCKNGSGSPCSIHSAKSQALKPKKFKYKD-----RTACWL 852
QY 296 -----VAHK-----VPGVT--DCVAEIQKNKPYIRHEPIGIDRSRRKYVFLNR 337
DB 853 KCMISFOLPAHRSRLSAILPFGMRGEAAAREKSALL--SPGSLRF----- 900
QY 338 LIIEDETFENNEKKIWIYSTKVOALALDCLDKQWEA--ELCKLLEEMREIHRHMOT 395
DB 901 -----TSTASSSPVY--VASTSVK-----WKSSGSTSALTISIKPNPLAENVT 943
QY 396 EDLTNKAAGS---NKSFL-----AAANEEILSEIRAK--KGDINVKSPETEEDK 441
DB 944 FGSPTLLRPLILLENPLFLKISNADQKLAALAISSPLTLKKNKQEKKEKYVKESEOKLL 1003

QY 442 NETENDSKD---AEKNREFEEDOSLEKSDDKTPDDDPQKSE-----EPTVEVD 489
DB 1004 SPTQAGTKKSGAEEAOVEEVPQKEAPQSTTTPPSASNGASHGVAQEAELAGENATGD 1063
QY 490 KONSVSANLGDWTTNATSEETSPSEGRSPVGLSE--TPDSSNMAEKKVASLEPODV-- 544
DB 1064 TLKRQRIDLPGRVNVHVCASAVILQF--PLATFGEQDQEPEDADMQEFLAALVPASIMEP 1122
QY 545 -PEEPNKTCESSTTSATTTSTIOPNLNENSSSELNNSQSESAKADDPENGERESHTPVS 603
DB 1123 SPEKPTHTIYTDENDNCASCSTSPVGEDSKPSKSGSAQAEVKA-----TA 1168
QY 604 IOEELVGDPTSEKSTGELSESPGAGKAGSGSTRITRLRNPDSKLSQLSQVAAAAHHA 663
DB 1169 LKE-----GTASAAGSSAKVY--GTN-----AAVAS 1194
QY 664 NKLTFEGKEVLVNVSGEISRLSTKEVKIMKNNYFRLGQEGKYRVYHNOYSTNFA 723
DB 1195 NLIVASKK-----QRNGDI--ATSSVTQSSN-----QTQGRKYEHRQORT---L 1236
QY 724 NKHQRHDH-----KRRHLA-----HKFCLT 745
DB 1237 ISIDWENDPRAEYQGTGRLVETVNAQRALCPLOGSTGLDPLIFCACCCEPHYQYVQ 1296
QY 746 PAGEFRKWSVHS--KVLITSLRLITTOLENNIPSEFHPNASHIRANVI----- 795
DB 1297 DEYNLK-----HGSFEDTTLMSLLETTVN--ASTGPSSSL--NOLTORLMLCPRCTVC 1347
QY 796 -----KAVQMSK-----PRFALALILBEAVKPYVMLPIW 827
DB 1348 YTCNMSSGSKVQCKQKQKYNHSTCLGTSKRLIGADRLICVNLCKKSCSTTKV----- 1401
QY 828 REFLG-----HTRL----- 836
DB 1402 SKFVGNLPMWCCTCFRLKRGKGNFCIQRCYDDNDPDKMMEGCGGQWHSCEGLSDEQ 1461
QY 837 HMTSTIEREEKKVKKKEKQOEELTMOQATWYKTFPYKHQVYKQK--GEERYVTGYGW 895
DB 1462 YNLSTLPESIEFICKKARNSS-----KIKAEWQAVAEERKASLYSVL 1509
QY 896 SWISTHYRFPYKLPQNTNVYRKSLBCTKNNMENDESDSKRCSRPKKIKTEP-- 952
DB 1510 KILSKSROACALLKLSPRKNV-----KCTGAGSSNGKGLQAPRAL 1548
QY 953 --DSEKEVYKDAKAGADQNMEDISKITEKKQDVKELLDSDDSKPCKEEMPEYDDMK 1010
DB 1549 QESSGSDNGLSD--GESQNSDDVYEFKDDQOQ--QQRANNMKP--RVKPL-----PCS 1598
QY 1011 TESHVNCQES--SOVDVNVVSEGFLRTSYKKTKSKSLDGLERRIKOFTLEEKQLEKI 1069
DB 1599 COQHTSHSQSFSLVDI-----KOKIAGNSYVSLLEBNYDMSQVIOO-- 1639
QY 1070 KLEGGIKGIGKSTMSKSNLSPYITKAKEGCQSDMKQBPSPANNQ-----PED 1122
DB 1640 -----SNCDLDAIYK-----ELLSEQFPWFQNETACTDALEED 1674
QY 1123 LIQGS-----QSDSVYLRMSDPSHTTNKLYPKDRYLDVDSIRSPTKCPKQNSIEND 1175
DB 1675 MFESGSGNYEEDLDQAGVSAVYNEHSTQSAESGYLD-----IPL----- 1717
QY 1176 IEEKVSDLASRQOEPTKSKTKGNDFFIDSKLASADIDGTLCKNK-----KPLI 1225
DB 1718 --EEVDLDGSGCIG--MRDPTWCLFCRKSGBGLSGEEARLLYCGHDQWHTNCAMSAEV 1774
QY 1226 QEESPTIYSSKSAHSSPVS-----TNDRDATPLSRANDFEGKLGCSDES-----N 1273
DB 1775 FEEDID--GSLQNVHSAVARGMTKCTVCGNGAT-----VGCNVRSGGEHYH 1819
QY 1274 STLENSDPTVSIDSS-----EEDMIYONSNSISQEFRTRODVYLEPLCELVS 1325
DB 1820 YPCASIDCAPLITDMSMCPAPAKGNALKANGSPSY-----TYSNFEVSPRYVEL-- 1872
QY 1326 GESTGNCEDRLPVKGTGANCKRPQOK--KLEER-----PVN-KCS 1363

```

1873 -----DRKKRLIE-----PARVOFHIGSLVRLQGAIVPRFSDSYEAVVPINFLCS 1919
QY 1364 -----DOIKAKNTDKKNENRESE-----KKGQRTSTFOJNGKDNKRTILKGCLE 1412
DB 1920 RLWSSKEPMKIVEYVTRTTIQNSSSTLALDVGRLVYDHNPNPSKEVQLMAQIARWH 1979
QY 1413 IESER---VSGNVE-----PKVNNINKIIPENDIKSLVKESAIPEFNGOYIMEDFNE 1464
DB 1980 TSLASEFLENGDMSGFEPPNNSC--VPPDONTBEEFOQADLPLPLKAIFEDL--- 2035
QY 1465 RNSSETKSHLSSDAEGN---YRDSLETLPSRKES--DSTQTTTPSASCPESSNVQVED 1520
DB 2036 -----PHELLDGISMLDIFLYDXTDLFAISEQSKDGTQAMTS-----NOAON 2078
QY 1521 MELETSVKVYKTSPTTSEESNLMDFTDENGILPI--KNENVNGSKRKKTIVTEYTTMT 1579
DB 2079 QNOQAAGANSVS---ICEDTRNNTSLIG--NGWPAASNVEEDAMLSAERN---SSQVOML 2130
QY 1580 STVA-----TESKTVIKVEKGDQOTVVSSTENCAKSTVTTTTLVTKLSTPSTGSDVD 1632
DB 2131 KTLAMPKLDGNSAMATAIKRRRLSK-----NLAEVFLILSSQQRKKKEATVAGV- 2181
QY 1633 IISVEQSTVVTVTYDLSLTGGTLVSMV--VSKEYSTND-----KVKLMKESRP 1683
DB 2182 -----SRROSISETSVGEVATTSGSVRSKFTWSAKRYFEKSEGREBAKRMIMQMDG- 2235
QY 1684 KTRRSTALPSYRKFTKSTKKSIFLPRMDLKLARKGIR---EVPYFNAPALPD 1739
DB 2236 -----VDDSIETFRITISGDNLSSTAOFSQOVQCDRCQCTYRINDAOR-- 2278
QY 1740 IMPYSPRPFTGLTWRYRQOTVKSAGVSLMLRLMLASLRMDMAKVPBGSGSTRTETS 1799
DB 2279 --HLBSCSPT-----MSSNETESDVS-----GGMTNNANO 2307
QY 1800 ETEITTEILIKRDDVGPVIGIREYCIRKTIICPIGVETPEKPTPOR-----KGLR 1850
DB 2308 ISAESINELKOLANAGLINV-----LOSATSPFOYRLGSLGOFGLIQ 2354
QY 1851 SSALRPK-----RPETPKOTGPVILETWAEELELM--EIRAFABEV----- 1891
DB 2355 QLOLOPOSIGNGEFLSOPNPATQANT-----DLOQIYANSLOGLANLGGGFTL 2403
QY 1892 -----EKEKAQAVEQAKKRLQOKPVTIATSTTSPTSTSTSPAKVAVAPI 1941
DB 2404 ACPVTYAPAPOLIAVSTNPDGTQCFIQILPQTMQATTTPTATYQTLQANTKIML--- 2460
QY 1942 SGSVTTGTMTVL--TKVSPATVTFQOKNFHOFATVVK-----QGSNSGVVOVQO- 1993
DB 2461 --PLTRAGKPLKTVATKAQAQAAVAKOROLKSGHYKPIQAKILOPHQHQOQOQTOVQOP 2518
QY 1994 -FVLG-----IIPSTGT-----SQQFTFSF----- 2013
DB 2519 ITVMGONLLOPOLLOSTSTQTOQAPQIILPQAPQNTIISVYTBGDSOGQPLQYISITPAGE 2578
QY 2014 ---QPR-TATVTRPNTSGSG-----TTSNS--QVITGPQIIRPMTVIR 2052
DB 2579 YKBPQPTATPFLTAPAGATYLOTDASGNVLVLTFTSNGSLQMLTQOSLOAQPOVIG 2638
QY 2053 TPLQOSTL-----GKAITRPVAVQAPAOVM-----TOITNG 2086
DB 2639 TLIQPTIOLGGGADGN-----QPSNOQPLILGTTGGSSGLEFATTSPOVILAT 2689
QY 2087 QVY-----STAVSAPNTVST--PG---OKSLTSATSTNLSIGSSASOPRPPQGOYK 2133
DB 2690 QPMYIGLEFIVQITVWSSQGFVSTAMPGLSONASFATTTVOFQSKIEP-----IYD 2743
QY 2134 LTMALQTLTQGHGNGOGLTV---VIOGOGO--TTGOLQILPQGVTVLPBGQOLMQAA 2187
DB 2744 LAGVYVLLNNTQDASAGFLNAASVLOQOQTODDPTTQI----- 2782
QY 2188 MNGVYORFLPLATATTTTSTVSTAAGTGEOROSKLSPOMOYHODKTLPPAOS 2247

```

```

DB 2783 LONANFO--FQSVPTSSGASTSMDYTSPVAVTA-----KIPPYQIKRTNAQAKAAG 2832
QY 2248 --SSVGAFAKAPQ-----TAQPSARPPQOTQPSAPQEPYQTOPEVQQTIVSSVBPSE 2299
DB 2833 ISGVGVPPPOPVVKNVLPSTISVTOOSQOVQVKNNSLQOSQVKGKAASGTTCGCA--PFSI 2891
QY 2300 A-----QPTHQSSKPOV-----AASQPSQNSVQOSQPVAV--Q 2331
DB 2892 ASKPIQKKTNTMIRPIHKLKVPKVKMPTPKYQONQNSILLQOQOQOQPOQOQILPAVVNQ 2951
QY 2332 SP-----SQTRIRPST-----PSQLSPGOQSOVQTTTSQPT---PIQHTISQ 2371
DB 2952 VPKVTISQORIPAQIQOQOQLOQAOMIHIPOQOQOPIQOQOQOVQPSMPTITLAEAPVQSO 3011
QY 2372 IPSQGOQPOQPOVQOSTORTLSSQOTLNOVSQSPSR----- 2407
DB 3012 FVMEPOLAEQOELANVQHFSTSSSSSSSSSSSCLPTNVVVPQOQAPSTISSSTRPTNRV 3071
QY 2408 -POLQIQOPQ-----QVAVPQOLOQOVQ--VLSQISQVNAQIQAOQSGVPOQIKLOLP 2459
DB 3072 LPMQORQEPAPLISNECPVVSPTPPKVPBQPIHQMTSASVSKYQAQSTLPSPV---YE 3128
QY 2460 IQLOQSSAVQTHQIQNVTVQAASVQOL-----QVQOQLRDQOQKKK 2502
DB 3129 AELKVSVALS--IYDVTMDAILQEPVTOISYTEGLYKNSPGESKTEOLLLOQOQRE 3186
QY 2503 Q--QOIEINVPNSKLLIVEITIKOVYMKHNAVIEH-----LKOKKS--MTPARBREN 2552
DB 3187 QLMQOLVNNGYLLDKHTFOVERPDYVYREDELEEDDEDDFSLKMRISACNDHEMSDS 3246
QY 2553 QRMIVCQNVKYYILDKIDKE-----KQAAKRRK 2581
DB 3247 EEPVAVKDKISK--ILDNLTDNDCADSIATATMEVDASAGYQQWVEDVILATTAQASAPTEE 3305
QY 2582 REBSVQKSKONATKLSLLKHKHQLAEILKRAALLD-KDLOLEVBELKRLDKTK 2640
DB 3306 FGCALETAAVEAAYIYNIMADAH-----VIDLKOLONGVELURR--KE 3349
QY 2641 EKDLMOAOTAAVAPCPVTPVLPAPAPPPSPPPPG 2679
DB 3350 EQRTVSQEBQSKALVP--TAAAPRPPPIQEPKMTG 3386

RESULT 3
K167_HUMAN
ID K167_HUMAN STANDARD: PRT: 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schlueter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO

```


OY	1402	KIYLKGECKLEISSRKVSQV--EPKVYNNIKLIPENDIKSLTVESAIRPTINDYME	1460
Db	1431	INAFRETAOKOLDPAASVTSKRHRPRTK--EKAKPLEDLAGW--KELFOTP-----VCT	1480
OY	1461	DENERNSEKTSHLSSSDAEAGVNRDLSLETLPSTKESDSTQTTTPSPASCPESNSVNOVED	1520
Db	1461	DKPPTHETTKIACRSPD-----PVDTPSSKPKGSKRLKRYD	1519
OY	1521	MELETSSEVKRYTSS-----PITSSEESNLSDPIDENGILPINK--NENVNGESKR	1568
Db	1520	VEEEFFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM--GTPQOKLDLTENLGSKR	1574
OY	1569	KTVLEVTYTMSTYA-----TSBKVIYIKERKDKOTVYVSTENCAKSVYTTT	1616
Db	1575	LOTPEKKAQALEDLAGEKLEFOTRHTHEESMT---NDKTAKAACKSSOPDLEDKNASSKR	1631
OY	1617	TYVYKLSPTSGSVSDIISYKESQKVVYTYVYDLSLTTTGTLVMSMYKESYPRDKVX	1676
Db	1632	RLKSLSG--KGVGEVEILLANGKLTQTSGETHTHTHPRTDGSKMAFMESPQOIIDSAKS	1669
OY	1677	LMKESRPKRTKSGTA-----LPSYRK--FVTKSTKKSIFVLPNDLU-----	1715
Db	1690	LTGSKROLRFPKSGKEVEDLAGEFIELFQPSHTKESMTNEKTXTVYSNAGOPDLVDRPT	1749
OY	1716	-----KKAARGGIRVPPYFNWNAKPAOLDIMPYSPRPRTFGITMYRIQIYKSLAGSL	1768
Db	1750	SSKPOPKRSUKRADTEE-----EFUAFRKQTPSAG---KAMHTPKPAVGEK	1793
OY	1770	MLR--LWASLMBDDMAKVPGGSGSTRJTESLTTTETIKIRRDVGPYGR-----	1820
Db	1794	DINFLEGTPOKLDQPOENLE--GSNRRIQTRKKKAQALEELT-----GPELEFOTPCT	1844
OY	1821	-----FEYCIKRIKC--PIGV--ETPEKETPTPOKGLRSS-----ALRPKR-----	1859
Db	1845	DNPTADEKTKKILCKSPQSDPADPTPTNKKQPKRSILKADVEEFILAFRLTPPSAGAM	1904
OY	1860	ETPQOTGPVILETWVAEELMEINAF--AERE-----KEAQAWE	1900
Db	1905	HTPAAA-----VGEK-----DINTFQVPEKIDLLGNLPGSKRRPQPTKEKAKALE	1952
OY	1901	QOAKRRLKQOPVIAISTSPSTSTTSPAPQKMWAPVIGSVYTGKWLITTKVSP	1960
Db	1953	DLAFKELFQTPGHTESMTDKITFVSOCKSPQDPDVKTP-----TSKQRLKSLGK-	2005
OY	1961	ATVTFQONKNFHOTFEATWVAKQSGNSGVVOQOKVL--GIPDSGTGSOQTFSTFQPTA	2018
Db	2006	-----VGVKEEVLVPVKEKLTQTSAGKTQOT-----	2028
OY	2019	TVTIRPNTSGSGGT-----TNSGOYI-----TGPOIRGMYVITRP-----LOQSTLG	2061
Db	2029	-----HRETAGDGSIKAFKFSANAKOMDPRANVGEMKRP-----RTPREBAOSTLEDLAGEF	2079
OY	2062	KAIRTPWVOPGAPQOQVOTIIRGQPVSTAVSAPMTVSSFPQOKSLTSAISTSNIOSSA	2121
Db	2080	KELFQTDPHHEESTTDDTKTKI-----ACKSPPESSMDPTST-----	2117
OY	2122	SQPRPOQGOVKLT--MAQTLTOLTGCHGNQGLTVYIGOCGOTFOLOLIP-----QGVTV	2175
Db	2118	RRREKTPIGKRDIYEELSALKOLTO-----THTTKVPQDEDEKGINV	2159
OY	2176	LPGHGOOLM-----QAMPNGVTO-----RFLF--PRLATATATASTTTTTV	2215
Db	2160	FRETAOKOLDPAASVTSKQOPTRPKGAKOPLIEDLAGLELQTFQTECDKPTTHEKTKTKI	2219
OY	2216	STTA-----AGTEB--OROSKLS--PQMOVHODKTLPEPAOSSSVGAKAKOPOTA-----	2260
Db	2220	ACRSPOQDPVGTPTPIERQPSKRSRLKADVEESLALRKRTBVGKAMOTPKPAGDEKDM	2279
OY	2261	-----OPSAKPOPOTOPSPAPKEVQOTOPVQOTQTVSS-----HWSEAOPT	2303
Db	2280	KAFMGTEVQOKLDLPGNLPGSKMRPQTPKPE-----KAQALEDLAGEKLEFOTQTECTQKPT	2332

QY	2304	HAQSSKPOVAASQPOSNVQOQSPVRQSPQRIPS-----TPS	2344	QY	2304	HAQSSKPOVAASQPOSNVQOQSPVRQSPQRIPS-----TPS	2344
Db	2333	-TDEKTKIACKS-PQ-----PDDVDYIPASTKQPKRNLRKADVEEFLAKRRTPSA	2388	Db	2333	-TDEKTKIACKS-PQ-----PDDVDYIPASTKQPKRNLRKADVEEFLAKRRTPSA	2388
QY	2345	-----QLSPGQSOQOVQTTTSSPIPIPIPHSL-OLP-SOGQPOQOP-----	2382	QY	2345	-----QLSPGQSOQOVQTTTSSPIPIPIPHSL-OLP-SOGQPOQOP-----	2382
Db	2384	GRADMTPEPNAVSDKINNTIETVEPIVQKLDLGNLRPSKRPQOPQPKKAALEDLVGFKEI	2443	Db	2384	GRADMTPEPNAVSDKINNTIETVEPIVQKLDLGNLRPSKRPQOPQPKKAALEDLVGFKEI	2443
QY	2383	-QVOSSQTOTLSSGGTLANOVSVSSPSRPOLOIQOPQVIAVP-----QLQOOVQVLSQI--	2433	QY	2383	-QVOSSQTOTLSSGGTLANOVSVSSPSRPOLOIQOPQVIAVP-----QLQOOVQVLSQI--	2433
Db	2444	FQTGHEHTESMTDKITEVSCSKSPQEPSEFKTSRSSQKRLIKPLVKVMKMEELVAISKLTR	2503	Db	2444	FQTGHEHTESMTDKITEVSCSKSPQEPSEFKTSRSSQKRLIKPLVKVMKMEELVAISKLTR	2503
QY	2436	QSOVAQAQAOQSOVPOQIKL--QLPQIQOQSSAVQTHQIQNVVQQAASVQEOQLQRYQ	2493	QY	2436	QSOVAQAQAOQSOVPOQIKL--QLPQIQOQSSAVQTHQIQNVVQQAASVQEOQLQRYQ	2493
Db	2504	TSGETTQTHHEPTSDSKSIKAFKSPQIILDPAA-----SVGSRROLTRFKKARA	2555	Db	2504	TSGETTQTHHEPTSDSKSIKAFKSPQIILDPAA-----SVGSRROLTRFKKARA	2555
QY	2494	LFDQOQKK-----QOOLEINVT-----PSKLLIKVEIIO-----KQVAKH	2531	QY	2494	LFDQOQKK-----QOOLEINVT-----PSKLLIKVEIIO-----KQVAKH	2531
Db	2556	LEDLVDFRELEFSAPGHTEESMTIDKNTKIKCSKPPRELIDTATSTKRCPTKREKVEE	2615	Db	2556	LEDLVDFRELEFSAPGHTEESMTIDKNTKIKCSKPPRELIDTATSTKRCPTKREKVEE	2615
QY	2532	NAVLEHLKQKKSMT-----PAERENORMIYCNQVMKYILDKIDKEEQAAKKRREES	2585	QY	2532	NAVLEHLKQKKSMT-----PAERENORMIYCNQVMKYILDKIDKEEQAAKKRREES	2585
Db	2616	LSAVERLTQTSQOSTHTHKKEPASGDEGIKVL-----KQAKKKRPVE	2658	Db	2616	LSAVERLTQTSQOSTHTHKKEPASGDEGIKVL-----KQAKKKRPVE	2658
QY	2586	VEQKRSKQNA-----TKISALLFKHKQOLRA-----ELKKRA	2618	QY	2586	VEQKRSKQNA-----TKISALLFKHKQOLRA-----ELKKRA	2618
Db	2659	EEPSRRRAKKEKAPLEDLAGTELESTSGHTQESTLTAQATKIPCESPLEVVDTTA	2718	Db	2659	EEPSRRRAKKEKAPLEDLAGTELESTSGHTQESTLTAQATKIPCESPLEVVDTTA	2718
QY	2619	LLDKDLQEOVEELKRDLIKKEKMDLQALQ-----ATAV	2653	QY	2619	LLDKDLQEOVEELKRDLIKKEKMDLQALQ-----ATAV	2653
Db	2719	STKHLKTRVQK-----VQKKEPSAVKLTQTSGETTDADKREAGDGKALKKSAQT	2773	Db	2719	STKHLKTRVQK-----VQKKEPSAVKLTQTSGETTDADKREAGDGKALKKSAQT	2773
QY	2654	AAPCPPTVPLAPAPPPSP-----PPPGVQHTGLSTPLLVAASQKRKREEKD	2705	QY	2654	AAPCPPTVPLAPAPPPSP-----PPPGVQHTGLSTPLLVAASQKRKREEKD	2705
Db	2774	PAPASVYTGSRRRAPRARESAQALIEDLAGKPDAAHTEESMTDKTKIKPCSSPLEED	2833	Db	2774	PAPASVYTGSRRRAPRARESAQALIEDLAGKPDAAHTEESMTDKTKIKPCSSPLEED	2833
QY	2706	SSSKSKKKMMSTSKETKTDKTLXCCKPRYDESKFYIGCBQCNWNGRCVGLGQSEA	2765	QY	2706	SSSKSKKKMMSTSKETKTDKTLXCCKPRYDESKFYIGCBQCNWNGRCVGLGQSEA	2765
Db	2834	TATSSKKRPRTRAQKVEKEE--LLAVGK-----LTQTSG	2866	Db	2834	TATSSKKRPRTRAQKVEKEE--LLAVGK-----LTQTSG	2866
QY	2766	ELIDIEYVQPCQSTEDAMTVLPTLEKDEYGLKRVLRSLQAHKAMPLEVPDPN-DAPD	2824	QY	2766	ELIDIEYVQPCQSTEDAMTVLPTLEKDEYGLKRVLRSLQAHKAMPLEVPDPN-DAPD	2824
Db	2867	E-----THTHDEPVEGSKGKA-----FKQAPKRVADAD	2897	Db	2867	E-----THTHDEPVEGSKGKA-----FKQAPKRVADAD	2897
QY	2825	YVGVKEBPM-----DLATMEERVQR-RYEKLETFVAD	2856	QY	2825	YVGVKEBPM-----DLATMEERVQR-RYEKLETFVAD	2856
Db	2898	VIGSRQRPAPRAKERAQPLEDLASFOELSCPQHTHEELANGAAD	2940	Db	2898	VIGSRQRPAPRAKERAQPLEDLASFOELSCPQHTHEELANGAAD	2940
RESULT 4							
ANK2_HUMAN	ANK2_HUMAN	STANDARD;	PRT; 3924 AA.	ANK2_HUMAN	ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
AC	Q01484; Q01485;			AC	Q01484; Q01485;		
DT	01-APR-1993 (Rel. 25, Created)			DT	01-APR-1993 (Rel. 25, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)			DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)			DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).			DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).		
GN	ANK2.			GN	ANK2.		
OS	Homo sapiens (Human).			OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;			OX	NCBI_TaxID=9606;		
RN	[1]			RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).			RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).		
RC	TISSUE=Brain stem;			RC	TISSUE=Brain stem;		
RX	MEDLINE=91302466; PubMed=1830053;			RX	MEDLINE=91302466; PubMed=1830053;		
RA	Ototo E., Kunimoto M., McLaughlin T., Bennett V.;		</				

Db 1778 PSLKSEBHAAGSPSPKTERHSTLSSAKTERBHPVPSPSKTEKHSPVSPAKTERHSPAS 1837
QY 616 KSTGELSESEFGAGKAGSGSTRITRLRNDPSKLSQLKSOQVAAAHAANKLREGEVLV 675
Db 1838 SSSKTEKHSP-----VSPSTR--TERHSPVSTKTERHHPVSPSG-----KTDKRPV 1883
QY 676 VNSQGEISRLSTKEVIMGNINNYFKLQCEGYRVYHNQYTNPSALNKHQRHEDHDC- 734
Db 1884 -----SPSGTEKHPPVSEKTEKRLVPSPSGR-----TDKHQVSTAGCT 1924
QY 735 RRLHAHKFCLTPAGEFKMWSVHGSKVLITSLRLITTOLENNIPSSFFHPNMAHRANV 794
Db 1925 EKHLP-----VSPSGTEKOPVSP-----TSKTERIETMSVREL----- 1960
QY 795 IKAVQMSKREFALALALIECAVKPVMLPIREFLGHRLHRMSTIEREEKYK-KK 853
Db 1961 MKAFOGSOODPSKHRTGTFEHRSAKOK-----OPOEGKAVKEV 1998
QY 854 EK-----KOEETMOQATVWKYTFPVKHOVMKQGEYKVTGSGMSWISKTHVRFVY 908
Db 1999 EKGPIILQRAQKTEMT-----KRGORLPVYG----- 2027
QY 909 KLPGNTNVTNRKSLBEGTKNNMDESDKRCRSRPFKIKTEPDSEKDEKSDAKKA 968
Db 2028 --TAESKRGVRSISGVKK--EDAGAQRKQVLSHKIPEVQSVPEBESHRESEVPEKKA 2083
QY 969 D-QNEMDISKITEKK--DOVKELLDSDDKPCKEPEVDDDM-KTESHVNCOESSQVDV 1025
Db 2084 DEQGDMDLQISPRKSTSTDESEVIAKQELENDKYOQFRLSEETEKQALHD-----QVLT 2138
QY 1026 VNVSEGFHLRTSYK-----KTKSSKLDGLE-----RIKQFTLEF 1062
Db 2139 SPFWTTPPL--DYMKDEFALSLQSGALDGSSESLKNEGVAAGSPGSLMEGTPOLISSE 2196
QY 1063 KORLEKIKLGGIKGIGKIGKJGKJGKJGKJGKJGKJGKJGKJGKJGKJGKJGKJGKJG 1119
Db 2197 SYKHE-----GLAETETSPESLFSFSPKSEBOTGKTESKJGKJGKJGKJGKJGKJGKJG 2247
QY 1120 -PEDLIGCSQSDSVYARMDSHTTNKLYPKRVLDDVARSPEKCPKONSLEND--- 1175
Db 2248 TYKOTGSGSERGATV--TEDSETSPESQKKEATLG--SPKQSPR--QDDCTG 2296
QY 1176 -----IEEVSDLASR--GQEPYKSKTGKNDFFLDDSKLASADITGLICK 1219
Db 2297 SCVALAKETPTGLTEBAACDEGORTFGSSAHKTQT-----DSEAQES--TATSD 2344
QY 1220 NKRKLIOEEDDTLYSSSKSLHSSVPKSTNDRAATPLSRAMDEKKGCSSESSTLENS 1279
Db 2345 ETKALPLPEASVKTDTGTESPOGAVIRSPQGLELALPSR-----DSEVLSAVAD 2394
QY 1280 SDTVSIODSSEEDMIVONSNESISEOPRTREODVEVLEPLKCELVSGESTGNCEDRL-- 1336
Db 2395 SLAVSHHDSLEASPVLEDNS--HKPDPSLP-----SPLKSPCDSLESS 2439
QY 1337 PVKGTENGKKPS-----OQKLEERVKNCS 1364
Db 2440 PVEKMRAGITFPHPLPAVAKTELLTEVAVSARLLRDPDGAEDSDLEQTSIMGSSG 2499
QY 1365 QIKLKNTDKNNRESEKKGQRTSTFOJNGKNDKPKYILKGELEKIEISERVASGNVE 1424
Db 2500 KSPILSPPTPSEEVSYETVKTIVST-----PKPAVHHCACEDD--SENGE 2545
QY 1425 PKVNNINKIIPENDIKSLTYKESAIRPFLNGDVIMEDFENERSSETSKSHLSSDABGNY 1484
Db 2546 KK-----RFPPEEMFMKVTIKIMFDELEQAKOKRDYKREPKQESS--SSSDPDAD 2597
QY 1485 RDSLETLPSTKESDSTOTTPSACSPESNSVNOVEDMEIETSEVKKYS----- 1533
Db 2598 SVYDVE-PKHTGSEDESGVPLVLTSESRYSSSESEPELAQLKKADSGLLPEPVIRV 2656
QY 1534 ---SPITSEESNLSNFI-----DENGLPINKENNVNGES--KRKT 1570
Db 2657 QPSPPLSSMDSNSSPEEVOFOPVYVSKOYTTKMNEDTOEBFGKSEEEKDSHLEADRHA 2716

QY 1571 VITEVTTMTSTVATESKTEVIVKEKGDQTVVSTENCASKSTVTTTTVTKLSPSTGGS 1630
Db 2717 VSTE-----ADRSYDKLNROTDPKIODGHGCEAMSSSARVSSGLSPGDD 2767
QY 1631 VDIISVKEQS--KTVVTTVTYDLSLTGGLVTSMT-----VSKYSTRODKYKLMKF 1680
Db 2768 VD-----EQPVLYKESIALOGTHEKDEGEELDVRAESPQADCPSESFSSSLPHCLV 2822
QY 1681 SRPKTRSG-TALPVRKRVTKSKSIFLFPNDLKLARKGGLREVYFYNAKPALD 1739
Db 2823 SEGKELDEISATSSIQKTEVTKTDETFENLPK----- 2856
QY 1740 IMPYSPRPTEFTWYRLOQTVKSLAGVSLMLRLMASLRLWDMAMAKVPPGGSTRTETS 1799
Db 2857 ---CPQDSSIT--QT-----DRFSMDVPVSDLAENDELY 2887
QY 1800 EPEITTTTELKRRDQPVYIREYCIKTLICIGV-----ETPKETPPQKGLRS 1851
Db 2888 DPQITS-----PYE-----NVPSSQSFSSSESKTQTDANHTTSEHS 2923
QY 1852 SALRPKRPEP-----KOTGPVLIETWAAEELELMEIRAFARVEKEKAQAVEQA 1903
Db 2924 SEYVSTITISPEVDVYVASSSGTVLSK-----ESNPEGDIK-----MESOL 2966
QY 1904 KRL--BQKRPVYATSTSPSTSTISPAQVMVAPISGVTTGKMYLT--TVGSPA 1961
Db 2967 ESTLWEMQDSV--SSSEPTMSATTVVGEQI-----SKVILTKTVDSQD- 3010
QY 1962 TYTFQONKRFHQTFATVWVQGSNSGVNOQKYLGIIPSTGTSQGTPTSPPTATVT 2021
Db 3011 --SWSEIRDDDEAFEARV-----BEOKITGLM-----VDRQ----- 3041
QY 2022 IRPNTSGSGTTSNSQVITGPQIRPGMTVIRTPLOQSTLGAALIRTPVWQDAPQOVMT 2081
Db 3042 -----SGGTT-----PDTTPARTPTEGCT--PTSEQVPLFLQSGK--LF 3076
QY 2082 QITRGQPVSTANAPNTVS--STPGOKSLTATSTISNIQASASOPPRQOQVQLTMAQ 2138
Db 3077 EMTRSGAIMTKRSYADESFHFQIOESREFTLSIDVKEGATGADPLE-----TSAE 3131
QY 2139 LQVLQGHGNGOGLVYVIGOGQGTGOLQLOIPQV--VLPBGQOQLMAAMPNGTVQ 2194
Db 3132 SLALS-----BSKETVDEADLILPDSVEEVEIPASQALN----- 3168
QY 2195 RFLFTPLATATATSTTTTVSTTAAGTGEOROSKLSPOMOVHODKTLPPAOSSSVGP 2254
Db 3169 -----SQMGISASTETPTKCAVSQT-----KDLPTVQGTGDIPLS 3204
QY 2255 AOPQTAP--SARQPOQT-----POSAPQ--EVQOPEVQOT 2290
Db 3205 GVRQISCPDPSSEPAVOYQDLFTLRSVYSDRGDSDPSSEPQOKSVIETPTAPMENVPF 3264
QY 2291 TVS-SHVPSEAOPT-----HAOSKP-----QVA 2313
Db 3265 TESKSKIPVMTPTSTIPAPSAEYESVEDTLSSVDEENKADKAPKSKLPKYVPLQV 3324
QY 2314 AOSPOPSNVQOSPVAVQSPQRTIRRPSTPSQLSPQOQSOVOTTSOPIPIOPHTSLQIP 2373
Db 3325 EQQLSDLDTSVQKTVAPOGDMAIASIPDNRSK--SESDASLSKTKCPKVTSTRYETETE 3383
QY 2374 SOQOPQ-----SQPOVOSTQTLSSGQTLNOV--SVSSPSRPOLO----- 2411
Db 3384 SREARELESEEGATRPVILTSRLPVKRSRTSSCRGQSTPKSKKEHFDLYRNSTE 3443
QY 2412 -----IQOPQOYIAVQLOQOVVLQSOIQVYVAOIQAOQSGVPOQIRLOLPIQIOQSS 2466
Db 3444 FFEESIDPASKLYDRLOQSRREQEIVYDDESSSALAEVSIENLPVETESHVEDIFDR 3503
QY 2467 AVQTHQIONVT-----VQAAVQEOQLQVQ-----QLRQOQKQKQ 2503
Db 3504 PWDSEIETLIERLPDENGHDHAEQDQDEQERIEERLAVYADHIGFSWTELARELDFTEE 3563


```
QY 2504 QQTEIVNTPSKL-----LIKV-----EITOKOVYMKHNAVIEHL 2538
DB 3564 QIHQIIEPNPNISQDOXQYLKIMLERDGNHATDNLVECLTKINMDIVHMETNTEPL 3623
QY 2539 KOKKSTPAERENORMATONQVMIYLD-----KIDKEKQAKKKRKESEVOKRS-- 2591
DB 3624 QERISHSYAIEIOT-----ITLDHSGFSYLOEBELCTAOKHOKOEQAVSKSESET 3672
QY 2592 -----KONATKLSALLF--KHKEQLRAEILKKRALLDKDQ 2625
DB 3673 CDHPPIVSEEDISVGYSTFODGVKTEGDSSTALFPQTHKEGVQDDFGSKMDLREESS 3732
QY 2626 IEVQELKADLKIKKCKDMLQAQATAVAPCPVP-----VLPAPRAPPSPPPPPGV 2680
DB 3733 LEVQOEY-----FVTPPTSETSETKAMIVPSPSPSTPREVSTPAE 3773
QY 2681 QHTGLSTPT-----LPVASOKRKEEEKDSSKSKKMMISTSKETKDKTKLCLCK 2734
DB 3774 BEKLYIQTPTSSRGGSPPIQEPPESEHREESS--PKRTSLVYESADNQPET-----CE 3827
QY 2735 TPYDESKFYIGCDRCOMWYHRCVGILOSEALIDEYVCPQOSTEDAMTVLPLTEK-- 2792
DB 3828 RLDEDAFAFKGDMP-----IPPEVTEEEYI-----DEHGHTVAVKVTARKII 3871
QY 2793 ----DYEGLRKRVLRSTQAHKMAWPLEVPDNDADPYGVIK 2830
DB 3872 RRYVSEGTETEKEIINVQMPQ-----EPVNIIEGDDYSKVIK 3908

RESULT 5
TRX_DROVI STANDARD; PRT; 3828 AA.
AC Q24742;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7244;
RN [1]
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RL Mech. Dev. 53:113-122(1995).";
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; 250038; CA690349.1; -
DR HSSP; P20393; 1A6Y.
DR Flybase; FBgn0014844; Dvir\trx.
DR InterPro; IPR003889; Fyricl.C.
DR InterPro; IPR003888; Fyricl.N.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
```

```
DR InterPro; IPR001628; Znf_Casteroid.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 3.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00184; RING; 2.
DR SMART; SM00317; SET; 1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 3.
DR PROSITE; PS50016; ZF_PHD_2; 3.
KW Transcription regulation; zinc-finger; metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator.
FT ZN_FING 1251 1334 PHD-type 1.
FT ZN_FING 1335 1380 PHD-type 2.
FT ZN_FING 1408 1469 PHD-type 3.
FT ZN_FING 1708 1767 PHD-type 4 (ATYPICAL).
FT ZN_FING 1768 1818 PHD-type 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-GLN.
FT DOMAIN 243 251 POLY-ALA.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 538 546 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
SQ SEQUENCE 3828 AA; 413721 MW; 32059CF303A3C504 CRC64;

Query Match
Best local similarity 18.6%; Score 407.5; DB 1; Length 3828;
Matches 691; Conservative 451; Mismatches 1260; Indels 1315; Gaps 177;

QY 5 EEEEDGDAEETDSEDEDEEEDDDSDVPEEM-----EDDDDDASYCT 51
DB 523 EDNDDGSA-----SSDAIEDDEDIDDDAEENEFAASEKSAETASVDEKEADQOLVM 577
QY 52 ESSEF--RSHTYSTGTGRKRPVHRRSPITLEKNDIPLEFPSSDDLWPNPHI-----M 105
DB 578 DKHFVLPKSTRSS-----RIIKPNKRILEVGGICSKRSPSDANGKPKKVFGLATL 630
QY 106 NVIAIEVLNFGTVL--RLSPRPEDFCALVSQRCITLAEHNVLLKAVLREEDTSN 163
DB 631 PAKCTRRRRRSATATLSQKIGKETFASFATAKNNS-----FVLNQPRQFOTDKS 681
QY 164 TTFGPADLKDSVNSTLYFLDGMTWPEVLRVYCESDKYHHVLP-----YQEAEDPYGPV 218
DB 682 RSEVSA--KPTLPPTT-----VLPASSAITSANVLSFGAL 715
QY 219 ENKIKVLOPLVDFLTNTNARELMSEGVYQYDDHRCVC-----HKGLDGLCC 266
DB 716 NN-----ANSVAASAATCAVCSAPVNNKDAPLARYG--VIGC 751
QY 267 ETC-----SAVYHDECVPPLLEVP-----DEMOCEVCAHVGVGVDCAVETGKRP 315
DB 752 EYCKRKNRSKRTKSKLSTPHNSPSTSTAGSQGLKC-----ID----- 790
QY 316 YIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWIYSTKVQALIELDCLDKDQWBA 375
DB 791 -----GGNCS-----ILSLKSQLKKNFKKL--YKERCKACWMLKCIATQLPLA 830
QY 376 -----ELCKTL--EEMREIHRHMDITEDL-----TNKAR 403
DB 831 GHRSLRSLAILPASMRREVAAPKDKCPDLLSPATSLRFTAPTSSASSGTTIKWSSAETAV 890
```

```

QY 404 GSKSFLAANE-----ELLEIRAKGIDDNVKSPEETE----- 438
Db 891 NSIKSNPLAENNTYFGOTPLLRPAILEKPLFLKIGSDNKRKAKESKEKLGISLVPVSTSEAA 950
QY 439 -KUKNEENDSKDAEKNREFEEDOSLEKSDDKTPDDPEQGSSEB---TEVGDKNSV 494
Db 951 VADGKTRKKOKREKARELEAKPLSPNAKKTTEANTPETOKDEOPASPTTVTVSAASS 1010
QY 495 SANIGDNTNATSEETSPSEGRS----- 517
Db 1011 TSHTSAAATNSQLETTTEANANASAVPDLNKRQIDLGPRVKKHVCSASIVLGOPLATFG 1070
QY 518 -----PVCLSETPDSSNAEKKVASSELPODVEEPNKTCESSNTSAT--TTS 563
Db 1071 DEEBELAAEACGAPPTTTTTTTSPEVILKKKPSPOPMOMTIDENDCASCILPTTEATAE 1130
QY 564 IOBNLEN-SNSSLSESNSSOSESAAADDPBNGRESHTPVSIOETIYGDPTSEKSTGELS 622
Db 1131 AOPAVSVLESRSKSKSTOTTEAKK-----TPA-----TSGSSGKVT 1167
QY 623 ESPAGKAGASGSTRITIRLRNPDKLSQLKSQOAAAAHAHANKLFEGKEVLVYVNSQGEI 682
Db 1168 TRNATFTVTSVASSLVATKKORNIIEVSSSISSQAAATOSRRALAK-----EV 1215
QY 683 SRLSTKEVYIMKNINNY-----FKIQCEK-----Y 709
Db 1216 NRL---KALISIDFWEYDPAEVCOTGFLITVEYVQORALCFELCSTGIDPLIFCACC 1272
QY 710 RYVHNGYSTNSFALNKHQREDHDKRHLAKCLTPAGFEKNGSVHSGKVLITGLRL 769
Db 1273 EPHQYCVLDEIYL-KHSEED-----TLMT 1297
QY 770 TITOLENNITPSPFHPNMA---SHRANWI-----KAYOMCSK----- 803
Db 1298 SLETSNNACAIASATNTALNQTLQRLNMLCPRCTVCYCNMSSGSKVKQKQXNVHST 1357
QY 804 -----PREFALALILECAVKFVYVLPIMREPLGHTRLHRTMSIEREKY 860
Db 1358 CIGTSKRLLGADRLPILCVNCLCKSCATTKV-----SKEVGN-LPMCTAC-----FKL 1404
QY 851 KKK-----EKKOEBEE---TMOQA---TWY-----KY---TFPV----- 875
Db 1405 RKGNGNCPICQKCYDNDPDLKMECGDCNQVHSHKCEGLSDEQYMLSTLPSIEFICK 1464
QY 876 -----KHQVKK-KGEEYRVYGGWSMISTKTHVYRFVPLPGNTVNNYRSLE 923
Db 1465 KCARODVSNKKADEWRQAWMEEFKSIVLYKLKLSRQACALLT--SPRKNMCCSA 1522
QY 924 GTRNNNDENWDESDKRCSSPKKIKIEPDSEKDEVKGSADAKGADQONEMDISKTEKED 983
Db 1523 GAOP-----AKAHSQKLOPKALOFTYNGL-GSDGESQNSDDIYEFKEQ-- 1565
QY 984 QDVKELLDSDPKCKEPEPEVDDMKTESHVNCQESSQVQDVYVNBEGFLKRSYKKTK 1043
Db 1566 -----HSTNKPSTPVP-----CSCLOPLSQSPSFL-VDIKOKTA 1600
QY 1044 SSKLDELERRIKOFTLEEKORLEKIKLEGGIKIGIKTSTNSKNLSSESVITKAKEGCO 1103
Db 1601 SNAYVSLAE-----FNVDMSQVIOQ-----SNDELDIAK----- 1631
QY 1104 SDSMROESFNANNDQ-----PEDLIOGCSQSDSSVLRMSDPSTTKLYPK---DR 1152
Db 1632 --ELLSEOPFPQNETKACTDALDEEDMFESCGYEELKESPTTYAAHHHTASQARFGLDI 1689
QY 1153 VLDVVSIRBPETKCPKRONIENDIEKVDLASRGOEPLKSKTK-----GNDFITDSKL 1207
Db 1690 PLDDVDVDLG---GC---AVKTRLDTRVCLFCRKSSEGLSGEABRLLYGCHDCVWH--- 1738
QY 1208 ASADIGTLICKNKRPLIEESPTIVSSKSALHSVPKS-----TNDROAT----- 1254
Db 1739 -----INCAMSAEVEEID---GSLQVNHSAVANGRMKICVCGKRGATVCCNVK 1786
QY 1255 -----PLSRAMD---FEGKLGDSBSNTLE--NSSDTVSIODSSEEDMIVQNSNES 1301

```

```

Db 1787 SCGEHYHPCARTIDCAFILDKSMYCPAHARNLKAANGSPSYVESNFEV-----SRPV 1840
QY 1302 ISBQFTRQEDVLEPLKCEL-----VSGESGNCEDR-----LPYKGTAE---NG 1345
Db 1841 YLELERKK---KLIVPAKQFHIGSAVNRQLSIYPRPSDSEALVPIFLCSRLYWS 1897
QY 1346 KRPQOKKLEERP-----NKCSDQIKL---KNTRDKNNENRESERKQ----- 1386
Db 1898 KEP---WKIYEIVTRTIQSYSSSTLILDAGRNFYTDHNTPNCSLVQGLAQIARHSSL 1954
QY 1387 QRTSTQJNGKDNKPKIYLKGECLKEI SESRVYSGVNERVNNIKITPENDIKSLTYE 1446
Db 1955 ARSDLLDTQWAE-FPNSVYPAD---ENTBE-----EQ-QVADLLPEE--IKD-AIRE 1999
QY 1447 SAIRPFLNG---DVIM-EDFNERN-----SSETKSHLLSSQAGNGYSDSLSTPSTE 1496
Db 2000 DLPHELDIGISMIDIFMYEELGDKTELFAWSEOSKGTATTSQAG--ASVILCDEDTKN 2057
QY 1497 SDS-TQTTTPSASCPSNSVNYQ-----EDMEITSEV-KKVTSSP-----I 1536
Db 2058 SNSLNKHVLSNCTJASNVPDAMLCAARSSQSEKCGDVLKTDIAPTRSMPKLDDGSV 2117
QY 1537 TSEESNLSDFTDENGFLPINKNENNVGSKRTVTEVTYMTSTVATESKYIYKEKD 1596
Db 2118 AAFKRRRLSKINAE--GVLLSLNQ---RSKK-----EMATVAGITR-- 2153
QY 1597 KOTVSTENCASJVTYTTTTYTKLSTPSTGSDYDIISYKESK-----FVYTTT-VT 1649
Db 2154 KOSVCGSSELPAEGSATMRTKSP-----TWSAACLEKESREBPALKITMOMGDV 2206
QY 1650 DSLTT-----TGGLTVTSMTVSK-----EXSTRDKVK-----LMKESRPKRTS 1688
Db 2207 DSITEFRIIGSDQNLSTAQOTGVKCEQOCYRANVDSQRHLGSCPEMSTSSSESEFAT 2266
QY 1689 GTALBPYKRFYTKSTKSLFVLBPNDLKK-----LARKGIGREVPYNNYNAKPLDIW 1741
Db 2267 GTAQLAESL-----NELQOVALAATLSMTGGL--NY-----LQTSF 2302
QY 1742 PVPSPRPT---FGITWRRLQTVKSLAGVSL-----MURLMASL 1778
Db 2303 POYQNLATLQGFQVQGLQGLQTLQ-LQPOSIGLGFLOPMNAQATNSGNDVQLYANSL 2361
QY 1779 RMDMAKVPYGGGSTRTESETEITTEILIKRDVPGYIGIREYCIRKILCPIGVPEP 1838
Db 2362 Q--NLANL--GGGFLTLQPTNSTQAPOLI-ALSTNPGG-----TQ 2398
QY 1839 KETPTPQKGLNSSALR---PKRPETPKOTGPVYIIETVAEELELMETRAFEREYERK 1895
Db 2399 OFIOLPQSGNATPOLLOTAAPRLCNAITYOT---LOATNSDKKIVLLFLEA-GDPLQEVY 2453
QY 1896 AOAVEO-----OAKKRLEQOKTVALATSTPSTSTSTTISPAOKV 1936
Db 2454 TOAAQOATAAHOKOLKSGHVKPQIAKLQOQOQOQHOOHQOHOQOQOQOQOQOQO 2513
QY 1937 MYAPISGYVTGKAVLTLTKVGSPPATVTEQONKNFQOTFAWVKQOSGNSGVYQVQKVL 1996
Db 2514 QOPRITVAQHGGTQLQGLQNLQOP-QLRFQSNMQ-PQLOQLLPLQVQAN----- 2561
QY 1997 GIIPSSTGTSQO-----TFTSOPRATVITIRPNTSGSGTT----- 2033
Db 2562 -IISFTYGGGSONOPLQYISIPPTTNDFKPQQTTSPTFLTAGGATFLQTDASGNMLMT 2620
QY 2034 -----SNSOVITGPQIRPMYVIRPPLQOSTL---GKAILITPVWVOP----- 2073
Db 2621 TAPANSGLQMLTG-QLOTQPOVIGTLQPOLTLTGADGTAPAAQOPLLIGATGGGT 2679
QY 2074 -----GAPQOYWTQILIRQOPV-----STAASAPTVVSST-PG---OKSLASATST 2114
Db 2680 TGLEPATAQOVIILAT---QPMYIGLETTIVONTVWSSQOFPVSTAMPGVLSQSSFSATTT 2735
QY 2115 SNTQSSASOP---PRP-----OOGQVKLTMALQTLQJGHGNGOGLT 2153

```

```

Db 2736 QVQASKEIPVLDPAQVVLNNAVNASGNTSLQDSQATDADATQALLQ--NAGFOFOT 2794
QY 2154 VVIQGGQQTGGQLDLPQGV-----VLPSPGQ-QLMQAA 2187
Db 2795 TPTTSTQQTSTQYAPRLVYAKVPAQOMKRNMTNANKSPISVLSKQVQPPQDSQVNVKV 2854
QY 2188 MPMGTV-----QRFLETPLATTAFTASTTTT-----VSTTAAG--T 2222
Db 2855 LPPNVVQQQQQQQQQQQQQQQQQQQQQLAGNANLKLTSQFORQQQANMELKNQAAGQQT 2914
QY 2223 GEGRQSKLSPQMVHDKT--LPPAQQSSVGR--AKAQPTQAPSAARQPTQPSRA-- 2276
Db 2915 GSTGCAAPSTASKPLQKTKMLNLRPHKVEYKPKIMQAPKLATSAASMHQHQQSPAAI 2974
QY 2277 -----QPEVQTPQVQVQVQTVSSHNPSAPQTHAQSRRQVAAQSQPQ 2319
Db 2975 NQVAKVALLQRLAPARQPP--QOEPOEQRHLHQQQQQQQQQQQHQHQQQQQLSMFQ 3033
QY 2320 SNVQGSQPV-----RVQSPSQTRIRPSTPSQLSPQSQSQVQVQTTTSQPIR----- 2363
Db 3034 L-LRAQQPPIISIVNTAEPQATQFVIRPALQAOAPQLQEQSQSQQQQQQAPBQLINGKA 3092
QY 2364 --IQPHTSLQIPSQ--GQPOQSPQVQSTQTLSSGQTLNQVSVSPSR-----PQLQIQ 2414
Db 3093 ARLQRYASNSLPTNVVNPLOQQRCASANNSSNVQONSTTINSRPTNVLPMQOQRE 3152
QY 2415 PGP-----QVIAVP-----QLQQQVQVLSQSQVY---AQIQAOQSGVPPQIKLQLP1-Q1 2462
Db 3153 PPLMSNDVYVQSPTPPKPIEPFYPAGASQYQKPIYKCYAQLEQKSPCYETELKTNITLNL 3212
QY 2463 QQSSAVQTHQIQ-----NVYVQVQASVQEQLOL-----VQQLRQQQQKKKQOQ 2505
Db 3213 EQNTSITTMQLOQPOQGPPIGEQIFEQSEAGVQLEKPKHNDMLLEATSCQQQQQQQQH 3272
QY 2506 IETNVMTPSKLILKVELIQOVYMKHNAVIE-----HLKQKS-----MT 2545
Db 3273 MEVNVNGFQVLSN-----ESCLLEKHFNVYEAVPMDTEHYASKMGSGGAAGIGQVD 3328
QY 2546 PAEREN-----QRMIVCN-----QVMKYLIDKIDKE--KOAAKKRR 2582
Db 3329 DABEDDDDDDFSLKMATSNACNDHMSDSEPAVKEKISILDNLTNDDQSDSIAVATTV 3388
QY 2583 ESEVEQRKSKQN--ATKLSALFKHKEQLRA-----ELIKRALLDKDLQIE 2627
Db 3389 EASAGYQOMVEDVLTATAGSVSTDETFATAEAVEAASYINEMAEMHQLKQLOAG 3448
QY 2628 VQBELRDLKIKKEXDLMQAOATAVAAPCPPTVLPAPAPRPPSPRPGVQHTGLS 2687
Db 3449 VELDLK-----KPKLDVPPQOQDPTVP--PNVVPYTAAPQOPPMRDPK-----IS 3492
QY 2688 TPTLPVQASQKRREKEDSSSKKKKMISTSKETKDKTKLCICTPYDES----- 2740
Db 3493 GPHLLKEIQSEQGYTKSSSIATIMEKVFPAVQVARRAH-----LTPLEPGLADMSG 3546
QY 2741 -----KFTY-----GCDRCQNMV-----HGRVGV-----ILQSE 2764
Db 3547 VQMIKLTAKNLKYLIEQLGVEKCVKTPKYHKRGNVSTAAGGHARTGSPAPALAG 3606
QY 2765 AE-LID-----EYVCQQCQSTEDAMVLTPLTEKDYGLKRVLSLQAHKMAWP 2812
Db 3607 ASSLIDVGSDDQELQENAEACARCE-----PYVSSEYDMFSWLSRRHKOPTOV 3656
QY 2813 FLEVPVNDAP-----DYGVGIKEPM-----DL----- 2835
Db 3657 FYQPSNENLVPRGCTGNSLPMAMKYRTIKETTYDYGVRSHIHGGLCTKRIEAGEMY 3716
QY 2836 -----ATNEEVQQRNYE-----KITE--FVADMTKIFNCRARYNSDSP 2873
Db 3717 IEYAGELLIRSTLTDK--RERYDSRGIGCYMFKIDDNLVVDATWRGAARFINHSCBP 3772

```

```

ID ATRX_HUMAN STANDARD; PRT; 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
RP VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Colleaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=874112;
RA Strayton C.L., Dabovic B., Gullisano M., Geetz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Gonzalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Geetz J., Pollard H., Gonzalez G., Villard L., Strayton C.L.,
RA Villasseau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKI1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
RA Colleaux L.;
RT "Specific interaction between the XNP/ATRX gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombor A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at

```

RESULT 6
ATTRX_HUMAN

Db 161 GEGJLHGT-----VSTACGQGVNH---FQKDSITRHPSLQ--VLICKNCEKYYM 205
 QY 234 TTNIABEELMSEGVIOYDHCRCVCHKLGLDLCETCSAVHYLCV-----KPLEEVED 288
 Db 206 SDDISND---SDGM---DECCRCWACGAGNLICDPCFNACCKCILRLNLRKELSTIMDE 259
 QY 289 --EMOCEVCAHVP---GVTC--VAE-----IQKNRPYIRHEPIGYDRSRKRYWELNR 336
 Db 260 NQOMVYIC--HPEPLDITVACNSVFENLEQLQOKKKIK---VDSEKSNKYEHTSR 314
 QY 337 RLIIEEDJE--NENEKTI-----WYSTKQVLAELIDCLDKDWEALCLILEMRE 366
 Db 315 FSPKKTSSNCGEKKLDSCSGSVTYSALLVPKEMI-----KRAKKLIETJAN 365
 QY 387 EIHRRHDIETLTKARGSN-----KSFLLAANEELIESIRAKKDJDN--VKSPE 435
 Db 366 MNSYVFKLQATDNGSISATKLRLQAKKSVLA---DIKKAHLALEDLINSFPAAMD 421
 QY 436 ETEKDKNETEN---DSKDAKNEDEEFDOSLEKSDDKT-----PDDDEQKSEEPTE 486
 Db 422 AVNKEKNTKEHKYIDAKFETKARKGEKPCALEKKDISKSAKLRSKQVDSEIHHQNVPT 481
 QY 487 VQKGNVSANLNDNTNATSEETSPSEGR--SPVGCLETFPDDSNMAEKVASELPQDVP 545
 Db 482 -----BORTKSTGGEHKKSDRKEEPQYEPANTSELDMDIVSPSSVPEDIF 529
 QY 546 EEBNKTC-----SSNTSATTTSTIOPNLNENSSSELTSSQESAKAADPENEGRE 597
 Db 530 ENETAMEVOSSVDHOGDSSGTBOEVESSVYKLNISKNRGIKSKITA---KYTKE 565
 QY 598 SH---TPVSTIQ-----EELVGFSTSEKSTG--ELSESPAGKAGSGSTRIITR-- 640
 Db 586 LYLKLPVSLPNSPIKAGADQOEVPQDKDYGSCGLNPKLEKCGIGQNSDNEHLVENEVS 645
 QY 641 -----LBNPSTKISQLK-----SQOVAANHAENKLEKEGEVLYV-----NSQ 679
 Db 646 LLIEESDLRSRVPYKTPTRPETETNPVTSNDECEETVKEKOKLSVPRKKDKRNSD 705
 QY 680 GEISR-----LSTKKEVIMKGNINNYFKLQEGKYRVH-----NOYSTNS---F 721
 Db 706 SALDNKRPKLPKSKOSEIYDONSDEDMALILKGYRMSHSSSDPDINDIEHTNHTLY 765
 QY 722 ALNKHQREDHDKRHLAHKFCILTPAGEF--KNGSVHSGKVLITSLRLTLITOLENNIP 779
 Db 766 DLTKQAGKDKGRKRRKSS---TSGSDFTKKGRKSAKSS---IISKRRQROSESS-- 815
 QY 780 SSFFHPRMASHRANMIKAYOMCSKPREFALALILECAVVPVMLPIWREFLGHTLRHM 839
 Db 816 -----NYSELEKEIKSMKIGAAR-----TTKKRIIPNT 844
 QY 840 TSIREEKEKVKKK-----EKKOBEETMOQATWVKYTFPVKHQ 878
 Db 845 KDDSDSEDEKSHKKGMDNOGHKMLKTSQESSSDAERKQORE-----TF----- 888
 QY 879 VMKQKEEYRVTVGYGWSWISKTHVYRVPKLPNGNTNVNRYKSLSEGTKNMNMENDES 938
 Db 889 -----SSAEGTVDDOTIMELRD-- 906
 QY 939 RKCSRPKTIKIPDESEKDEKVSDAKADON--EMDISKITEKKQD-----VKEL 969
 Db 907 ---RLPKQOQASATD---GVDLKSGKQSFSTLEVRKVAETLEKKSHLTKTKCKY 957
 QY 990 LQSDSDKP---CKEPEMEVDMDKTESHYNCOESSQVDVAVNSEGHLYTSYKKK--- 1041
 Db 958 QDLSIDIAEFLKDKDSDSEI SEDKSKKGTBEKKK-----PSDRKKVYIKM 1005
 QY 1042 ---TKSSKLDILR-----RIKQFTLEKQRLERIKLEGGIK----- 1076
 Db 1006 EQQYESSDQTEKLPREHEICHFPKGIKQIKNGTJDEGKSKKIRKTSKKKDELSDYAE 1065
 QY 1077 ---GIGKTSSNKKLSSESVYITKAKEGCO--SDSMQOESPPANNDQPEDLIQGCQSD 1131
 Db 1066 KSTGKSDSCSDSKSKNGAYGREKKRCKLGLKSSRRKROD-----CSSSD 1111

QY 1132 SYLVRMSDPSTHTNKLVPKDRVLVDVYSIRSPETKCPKONSIE-----NDIEKVSOLA 1184
 Db 1112 TEKYSMEKDCNCS-----DKRLKRIELRERRMLSSKRNKKEILOSQSSSDAEBSDDNK 1166
 QY 1185 SRGOEPTKSKTKGNDFFIDSKLASADDITGLICKNKKPLIOESDTIYSSKSALHSSV 1244
 Db 1167 KKKQR--TSSKKK-----AVIYKEK-----RNSLTST 1193
 QY 1245 PKSTNDATPLBRAMDEEKLGDGDSNSTLNSSTVSIOD-----SSEEDMIYON 1297
 Db 1194 KKKQADITS---SSSDIE-----DDQNSIGRESSDOKIKPYTENLVLSHTGFCQSS 1245
 QY 1298 SNESISEQFRTRDOVE-----VLPLKCELYSGESTGNCEDRLPYKTEAN 1344
 Db 1246 GDEALSKSVAVTYDDDDDDNDPENRIAKKMLLEIKANLSSDDE--GSSDDE--PEEGKRT 1303
 QY 1345 GKRPQOKKLEB--RPVNCSDQ-----IKLKNTT-----DKKNENRE 1381
 Db 1304 GKONEENPGDEEAKNOVNSSESDSESKRPYRHRLLRKLTVSDGSGEKKTKPREH 1363
 QY 1382 SEKKGQRTSTFOJNGKONKPKIYKLGECLEISPSRVYSGNVPEKVNINKIIPENDIKS 1441
 Db 1364 KEVYGRNRR--KVSSEDSJSDFOESGVSEVSSB--EDQRPRTRSKAKALEENORS 1418
 QY 1442 LTVKESAIRPFINGDVIEMDFENRNSSEKSHLSSSDAGNRYDSLETLPSTRES--- 1497
 Db 1419 YKQKKRRRIKVGEDSSSE--NKSNSDEEKEKEEEDDEEDDNDOSKSPGKG 1476
 QY 1498 -----DSTQTTTPSACPSNSVNOVEDMEIETSEVKY-----TSSPITS- 1538
 Db 1477 RKRIRKILKDKLRTETONALKEEERRRKRIABERERERKLEIEDASPTKCPITTK 1536
 QY 1539 -----BEESNLSDNFDENCLPINKNENNG-----ESKRKT-----VITEVT 1576
 Db 1537 LVUDEDETEKPLVOYHRNVIKLPKQHDVQVOPMMDCCESVAKTIKSPSGCIIAHCM 1596
 QY 1577 TMTSTVATES--KTVIYKEKGRQTVY 1601
 Db 1597 GLGKTLQVSEFLHTVLLCDKLDPESTAL 1623

RESULT 7
 ANK3_HUMAN
 ID ANK3_HUMAN STANDARD; PRT: 4377 AA.
 AC 012955;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Ankyrin 3 (ANK-3) (Ankyrin G).
 GN ANK3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=95138209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.;
 RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized
 at the axonal initial segment and node of Ranvier.";
 RJ J. Biol. Chem. 270:2352-2359(1995).
 CC -!- FUNCTION: Membrane-cytoskeleton linker.
 CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 CC alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues.
 CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U13616; AAA64834.1; -
DR HSSP: P55273; 1B18.
DR Genew: HGNC:494; ANK3.
DR MIM: 600465; -
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000488; Death.
DR InterPro: IPR000906; ZUS.
DR Pfam: PF00023; ank; 24.
DR Pfam: PF00531; death; 1.
DR Pfam: PF00791; ZUS; 1.
DR PRINTS: PR01415; ANKYRIN.
DR SMART: SM00248; ANK; 21.
DR SMART: SM00005; DEATH; 1.
DR SMART: SM00218; ZUS; 1.
DR PROSITE: PS50088; ANK_REPEAT; 21.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat.

FT REPEAT 73 102 ANK 1.
FT REPEAT 106 135 ANK 2.
FT REPEAT 139 168 ANK 3.
FT REPEAT 172 201 ANK 4.
FT REPEAT 203 230 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 267 296 ANK 7.
FT REPEAT 300 329 ANK 8.
FT REPEAT 333 362 ANK 9.
FT REPEAT 366 395 ANK 10.
FT REPEAT 399 428 ANK 11.
FT REPEAT 432 461 ANK 12.
FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA; 480399 MW; FA2379E55768B684 CRC64.

Query Match 2.78; Score 406; DB 1; Length 4377;
Best Local Similarity 17.68; Pred. No. 3.9e-07;
Matches 558; Conservative 444; Mismatches 1118; Indels 1048; Gaps 126;

QY 17 QDSEDEDEDEDEDDDDSDY-----PEEMEDDDDDA----- 47
DB 1441 KETSSDDDELKTRKQSFASALARKRYSLTEPGMTERSTGATNSLPTTYISKPFST 1500
QY 48 ----SYCT-----ESSFRSHSTYSSTPGRRKP-----RVHRPSPLIEEKDIPPLE 89
DB 1501 RPYQSMTTATITVPGPAKSGFTSLSSSSSNTPSASPLKSIWVSTP-SPIKSTLGASTTS 1559
QY 90 FRPSSEDLMPNENHIMVIAIYELRNFTGVLRISPPREFDCAALVSOECYLMADMHV 149
DB 1560 SVYSISISVSP-----INSLRTMSSPIKTVVOSQSPYNIQVSSGTLARAPAVTEAPPLKG 1613
QY 150 VLLKAVLRREDTSNTTFG-----PADLKDSVNSTLKFIDGMPTEVLRYVCS 197
DB 1614 LAANSFSSSTSPYTATGSLIERSSTTTPPAPSKNTNK----- 1653
QY 198 DKLEYHNVLRYQ-----EAEDYRPGVVENKIKVLQFLVDOFLTTNIABEELMSEGVIOYD 253
DB 1654 ---YSSSLPFKSLITTSAPLISSPLAKSVSPVAKSRVDVSSAKITMASSISLSSVYKQMPGH 1710

QY 254 CRVCHKLGLD-----LCCEITCSAVYHLECVAPRPLEEVEDEWQCEVCAHKVPGV 304
DB 1711 AEVALVNSISPLKXASSSTLINGCKATATLD-----EKSSATNSVSSVSAT 1760
QY 305 DCVAETQKN-----KPYIRHEPIGYDNRNR-----YWFNLRLIIEEDTENENEK 350
DB 1761 DTVEKVFSTTAMPSPSLRYSVSAFOSILRTPSASALYTSLGSSISATTSVTSII 1820
QY 351 KIMVYST-----KVQALIELDCLDWDYEAELCKLLEEMREIIN--RMDITTEDLTNKAR 403
DB 1821 TVPYSVVNVLPBPALKLPDSNSFTKSAALISPIKTLTETHPQPHFSSTSPV---- 1876
QY 404 GSNKSFLLA-----ANAEILIESIRAKKDIDNVKS-----PEETE 438
DB 1877 -KSLFLAPSAKLKLISTPSSLSQOELKLDVAEMKEDLMRMATLQTDVPEEKPPQPELPK 1935
QY 439 KDKNETENDSKDAEKNRFEEDQS--LEKD--SDDKTPDDPEEQKSEPTVEGDKGNSV 494
DB 1936 EGRIDDEEPFKIVEKVEDLVKSEILKKDYCVDNK--GSPKSPKS-----DKGHP 1985
QY 495 SANIGDNTNATSEETSPSEGRSPVGCSETPDSSNAEKVASELPODVPEERN-KTCE 553
DB 1986 E---DDWTFSSSEET--REAR-----QQAASQSP-SLEPRVQVAKA 2022
QY 554 SSNTSATTTSIQPLNENSSSELSNQSQESAKAADPENEGERSHP-VSTQEE----- 607
DB 2023 ASEKDYNLTKYIDLTDNIGSSSLNLKRYKPEDAKKQEGQKVLVPAIQLQHKLMK 2082
QY 608 --IVGDTSEKSTGBELSESPGAKGASGSTRITRLRNP-----SKLS---QLK 652
DB 2083 PASMRSTSEKEKCKMADS-----FFGTDTI--LESPDDFQSHDDKSPLSDSGFE 2133
QY 653 SOQVAAAHAHA-----KLEKE-----GKEVLV-----NSGELSRLSTKKEVI 692
DB 2134 SEKTPSPASQAEETGPKPLFHEVPITPVYETETREVVHVRISYDPSGADVPOQPEEPVS 2193
QY 693 MKGNINNYFKLGQGGKYRVYHNQYSTNSFALNKHQHREDHDKRHLNKKCLTPAGEFKW 752
DB 2194 PKPS-PFMEL--EPKPTTSIKKVKAFQKAKSEEDHN--RVLS----- 2235
QY 753 NGSHGSKVLITSLRLITLOLENNIPSSFPHWASHRA-----NWKAVOMCSKR 805
DB 2236 ---KGMKRVKEETITTTTRVYVHSPGG-----EGASERIEETWSVDIKAKQSGRDP 2287
QY 806 EFALALAILBCAVKPVVMLPIWREFLGHTRLNHR--TSIEREKEKVKKKKEKQEEET 862
DB 2288 KELAGLEEHKSAYSVDY-----HKSAETSAQHAEKDKQMKPLERITE-- 2331
QY 863 MQQATWVKYTFPVKHNQYWKQGEERYVTGYGWSMISKTIVYRFPKLPQNTNVNFKSL 922
DB 2332 -----VHIEKGNQAEPTLV-----IIRTKKHPRKEWVYVQKDL 2365
QY 923 -EGTKNMDEMDESKRKSRSPPKKIKIEPDSKDEYVKSQDAKADQEMDISKTEK 961
DB 2366 SRGDIINKDLPEKHNDFPSE-----OGQOEEBELTAEESSIPYLSSESRVNTPVSOE 2419
QY 982 KQD--DVKEILDSDQPKC---EEMPEVDD---MKTSHVNCQOSSQVDVNVSEGF 1032
DB 2420 EDSRPSAQLISDPSYTKLLLSQHSIEYHDELSELRGSGSYFAEKMMLSEKLDS-- 2476
QY 1033 HLRTSYK---KTKSKSLDGLLER-RIKQFTLEKQRLKRIEGLIKIGTSTNSSK 1087
DB 2477 HSDTEESVTDHAPRSELOGSDKRSRKHATAPKKEIISKIKYKDVSENGVAKVS----- 2531
QY 1088 NLSSEPVITAKGCGQSDSKROQSPNANDQPEDLIQCGSQSDSSVLRMSDPSHT-- 1143
DB 2532 -----KDEHFDDVTVLHYSGN-----VSSPKHAMMMR 2558
QY 1144 -----TNKLYPKRDVLDVYSIRSPETCKKONSIEMLDEKYSDLASRQGE--TKS 1193
DB 2559 FTEDRLDRGHEKLITIEDRV--DRTVKAEEKLTVEVSQFFRDKTEKIND--ELQSPKKA 2613

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE-87194600; PubMed-1106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STR1.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN-SPX101-1C;
RX MEDLINE-89031230; PubMed-1141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the StrA and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 238061; CA86176.1; -;
DR EMBL: M16164; AA35014.1; -;
DR EMBL: M16165; AA35015.1; -;
DR EMBL: X13857; CA32069.1; -;
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MUC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH
FT CARBOHYD 817 817 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C0DE2DBD61AA9D CRC64;

Query Match 2.7%; Score 401.5; DB 1; Length 1367;
Best Local Similarity 21.2%; Pred. No. 1.5e-07;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;

QY 1261 DEFG---KLGDSF-----SNSTLENSSDTVSIODSSEEDMIVONSNEISIQFRT 1308
DB 191 DPFQYWNIDCDNNGCGTKSTSTTSSESSTTSSES-----TTSTTSSESSTT 243
QY 1309 REQVLEVLKCLVGESEGNCE--EDRLPVKGTENGKKRPSQOKLEEPVKKCSDOI 1366
DB 244 TTSSTSESTSSSTTAPAPPTTCTKRPPTTCTKRPPT-----PHDDTTPCT 296
QY 1367 KLTUTTKKNNRESEKGRSTFOJNGKDKPKIYLKCEKLEISESHVSGVNEPK 1426
DB 297 KKKTTTCTCKTTTTPPTSSSTTE-----SSAPVPTPSSSTTESSAPVTSSTTE 350
QY 1427 VNNINKIIPENDIKSLVKESEAIRPFINGDIVIMEDFNERNSSEFKSHLLSSDAEGNYRD 1486
DB 351 ---SSAPVP---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY 1487 SLETLPTKESDS---TOTTPSASCPESSNVNOVEDMEFTSEVKKVTSPITSE---- 1539
DB 395 VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV---TSSTTESSAPVTSSTTES 453
QY 1540 -----EESN--LSNDFIDENGLPIKKNENNVNGESKKRKYTEVTYTMST--VA 1583
DB 454 SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV 513

QY 1584 TESKTVIKVEGDKQVVSSTENCASKSTVTTTTVTYTKLSTPSTGSGVDILISVEQSKTV 1643
DB 514 TPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSS-----STTESSSTP 568
QY 1644 VTTTVDLSLTGTTGTLVSNVSKKE--YSTRDKYKLLKESRPKKTNSGTLPLSTRKVVK 1701
DB 569 VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS 628
QY 1702 STKSIIVLENDOLKILARRGGIREVPYFNWNAKPAIDIMYDPSRPTFCITWRRLQTV 1761
DB 629 STESSSA-----PVPTPSS-----TTE 647
QY 1762 KSLVAGSLMLRLMLASLRMDMAKVPGGGSTRKETSETEITT-----TELIRK 1811
DB 648 SSAPVPTP-----SSSTTESSAPVPTPSSST--TESSAPVTSSTTESSAPVTSSTTE 701
QY 1812 RDGPPYIREYCIKRIKICPIGV-----ETREKPTPORKGIRSSAL----- 1854
DB 702 SSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSS 761
QY 1855 ---RPKRPTPKOTGVITETWVAEELEMEIRAFAEVKEKKAQAVQOAKKRLREQ 1910
DB 762 APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSSNTTSS 821
QY 1911 KP--TVIATSTTS-----PTSSTTISPAOKVMVAPISGVTGCTKMYLTTKYGS-- 1959
DB 822 APSSTPSSSTTESSSVPTPSSSTTESS-----APVSSSTTESSVAPVPTPSSSSN 874
QY 1960 ---PATVTFQOKNKNHOFATVYKOGNSGVVQOAKVLG-----IIPS----- 2001
DB 875 ITSSAPSIPIPFSSSTTESFST--GTTVPSSSKRYPGSOJETSVSTTETTVPTKTTTSSVT 933
QY 2002 -----STGTSQ--OTFTSFOPRTAVTIRPTPSSGGTSSNSOYITPQIRPM 2048
DB 934 PSTTTTITTCSTGCTNAGETTSKSPKVTYTV--PTTTTSTVTSSTTTT-----T 965
QY 2049 TVIKTPLO---QSTLG--KAIIRT--PVWVQCAPQOVMTQIIRGQVSTAVS-----AP 2096
DB 966 TVCTSTGNSAGETTSKSPKVTITTVPCSTSPSESTSPPTPVTVTVSTTVTTE 1045
QY 2097 NTVSSSTGQSLNSATSTNSIQSS--ASOPRRPQGGV--KLTMALQLQLQNGHNGGLT 2153
DB 1046 YSTSTKSGE--ITVTEVTKNIPPTVLTIAPTBSVTVTVTFPTTITTVCSGTGNSA-- 1102
QY 2154 VVIQGGOTTGOLILPOGTVL---PCPGQOLQMAAMNGTVORFLPLATTATAS 2209
DB 1103 -----GETTSGCS--PKVITTVPCSTGCGEYTTTA-----TTIVTTTA----- 1138
QY 2210 TTTTSTTAAGTGEOROSKLSPOMOVHODKTLPPAOSSSVGPAAKAPOTAPSPARPQ 2269
DB 1139 -VTTTVVTTTSSGTGNSAGKTTTG--YTKKSVPTTVTVTLAPS----- 1178
QY 2270 TQPSAPAPVQVQPEVQVQTVTSNHPSEAOPTHQSSKPOVAQOSQPOSNNQOGSPVR 2329
DB 1179 ---APVTPATNAVPTTITTECSAATNAAGETTSVCSAKTIVSSASAGNTAPSAF--- 1231
QY 2330 VQSPQTRIRPST--PSQLSPGOQOVQTT--TSQPIPIQPTSLQIPSGQPOSQPV 2384
DB 1232 --PVTITAI--PTVITTESSVGNAGETTVGTYTKKIPITTYITTL--IPSNAKNNEYV 1287
QY 2385 QSSTQTLSSGOTLNOVSVSSPSRPQLOIQOPQOVIAVLPOLQOOVOYLSQIOSQVVAQIO 2444
DB 1288 ATAT-----NPISIKTTS-----QIATPASASVAPV 1315
QY 2445 AAGSGVPOQIKLOLPIQOOSAVQTHQIONV--VTQOAS 2483
DB 1316 TSPS-----LTGPILOSAGSAVATVSPISSTYOGAA 1348

RESULT 9
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC 002817; 014878;

FT	CONFLICT	1351	1351	H -> L (IN REF. 3).
FT	CONFLICT	1412	T -> S (IN REF. 3).	
FT	CONFLICT	1449	L -> P (IN REF. 3).	
FT	CONFLICT	1504	M -> T (IN REF. 3).	
FT	CONFLICT	4192	G -> S (IN REF. 2).	
50	SEQUENCE	5179 AA;	85CD7571B9A5663 CRC64	

Query Match	2.6%;	Score 391;	DB 1;	Length 5179;
Best Local Similarity	22.5%;	Pred. No. 1.6e-06;		
Matches 221; Conservative	91;	Mismatches 418;	Indels 254;	Gaps 33

QY	1492	PSKSDSGQPTTSPASOCESNSVNOVEDMELETSVKKVSSPITSEESNLSNDFIDE	1551
Db	3460	PTTPTTTTTTYP-PTPTGTQPTPTTPTTTTTTYPPTPTGTQPTT-----	3508
QY	1552	NGLPINKNENV-----NGESKRTVTEVMTST---VATESKVIVKRGKQTVV	1601
Db	3509	-TPTTTTTTYPPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYPPT	3566
QY	1602	STENCASTYTTTTTATKLSPTSGSVLISVKEOSKVVTTVDSLTGCT--L	1659
Db	3567	PTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTPT	3626
QY	1660	VTSMVSEKSYSTRDKVKLMKFSRPKTRSGTALPSYKAFVETSKTSKSIIVLNDLKLTA	1719
Db	3627	TTTTTYP-----PTPTGTQPTTPTTTTTTYP-----	3654
QY	1720	RKGIREVPYENAKPALDIMPYSPRPFGITWRYRLQTVKSLAGVSLMLRLMASLR	1779
Db	3655	-----VPTPTPTGT-----QPTTPTPT-----	3673
QY	1780	MDMAAKYPPGGGSRTER-SETLETTTTELIRRDVGPGLGFECIRKILCIGP--E	1836
Db	3674	---TTTTTYPPTPTGTQPTTPTTTTTTYPPTPTPTPTPTPTPTPTPTPTPTPTPTPT	3720
QY	1837	TPKETPTQORKLRSSALRP-----KRPETPKOTG-----PVILETWAEELAMEI	1884
Db	3721	TTTTTYPPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYP-----	3771
QY	1885	RAFAERVEKKAQVEQAKRLEBOQPLVITSTTSPS---STTSTISP-----AQK	1935
Db	3772	-----PTPTPTGTPTPTPTTTTTTYPPTPTPTPTPTPTPTPTPTPTPTPTPTPT	3804
QY	1936	VWAPISGSVT-----TGKMWLETKGSPATYPOONKNPHOFATWVWOGOSNGV	1988
Db	3805	PTTPTTTTTTYPPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYP-----	3861
QY	1989	VOVOQKVLGIIPSSGTSGOFTSFQOPRTATVIRPNTSGSGTSSNSGVITGPQIRPM	2048
Db	3862	-----VPTPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTPTPTPTPTPTPTPTPTPT	3908
QY	2049	TVIRTPLOOSTGKALIRPVAVOGAGPQVWQIIRGQPVSTANASAPNTVSSPTROKSL	2108
Db	3909	TPPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYPPTPTPTPTPTPTPTPTPT	3963
QY	2109	TSATSTNIOSSASOP-PPPOGQVKLFMAOLTOLOGHGMGLVTVIOGOGQUTGQLO	2167
Db	3964	QPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYP-----VPTPTPTGTQPTTPT	4015
QY	2168	LIPGVTVLPGGQOLMOAMPNGTVQRFLEPLATTT-----ATTASTTT	2212
Db	4016	PTTTTTTYPPT	4066
QY	2213	TTVSTTAAGTGQRQSKSPQWQVODKTLPPAOGSSVGPAPAKAQOTAPSPAPQOTOP	2272
Db	4067	TTTTTYPPTPTGTQPTTPT	4114
QY	2273	OSPAPQVOTQEOVQOTQTVSSHVSEAOPTNAOSSKPOVAAOSQPOQSVWQOSPPRVOS	2332
Db	4115	VTPPTPTPTGTQPTTPTTTTTTYPPTPTGTQPTTPTTTTTTYPPTPTPTPTPTPTPT	4172
QY	2333	PSQTRIRS---TPSOLSPGOOSOVQTTTSPPLPQPHNTSLQIPSGOQO-----SQPOV	2384

Db 4173 PTTPTPTTTTTTVPPTPTGTCGCGPPTHTSTADIALETTSNPPESPSTQTSKTSSELT 4233

QY 2385 QSSQTGLSSGQTLNQVSSPSRP 2408

Db 4233 ESTTLSTLTPPAIEMTSTAPPSTP 4256

RESULT 10

AKA9_HUMAN	STANDARD:	PTT: 3911 AA.
AC Q99996;	Q9U004; Q9U0H3; Q9Y6Y2; O14869; O43355; O94895; Q9Y6B8;	
DT 16-OCT-2001 (Rel. 40, Created)		
DT 16-OCT-2001 (Rel. 40, Last sequence update)		
DT 16-OCT-2001 (Rel. 40, Last annotation update)		
DE A-kinase anchor protein 9 (protein kinase A anchoring protein 9)		
DE (PRKA) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)		
DE (Hyperion protein) (Yotlio protein) (centrosome- and golgi-localized PKN-associated protein) (CG-NAP).		
GN AKAP9 OR AKAP450 OR AKAP350 OR KIA0803.		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		
RP SEQUENCE FROM N.A. (ISOFORM 4).		
RC TISSUE=Brain;		
RX MEDLINE=98151389; PubMed=9482789;		
RA Lih J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M., Yotlio, a novel protein of neuromuscular junction and brain that interacts with specific splice variants of NMDA receptor subunit NR1.";		
RL J. Neurosci. 18:2017-2027(1998).		
RN [2]		
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.		
RC MEDLINE=99219864; PubMed=10202149;		
RX Witczak O., Skalihegg B.S., Keyser G., Bornens M., Tasken K., Janssen T., Oestlavik S.;		
RA "Cloning and characterization of a cDNA encoding an A-kinase anchoring protein located in the centrosome, AKAP450.";		
RL EMBO J. 18:1858-1868(1999).		
RN [3]		
RP SEQUENCE FROM N.A. (ISOFORM 3).		
RC TISSUE=Brain;		
RX MEDLINE=99287934; PubMed=10358086;		
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y. . .;		
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that anchors multiple signaling enzymes to centrosome and the golgi apparatus.";		
RL J. Biol. Chem. 274:17267-17274(1999).		
RN [4]		
RP SEQUENCE FROM N.A. (ISOFORM 1).		
RC Kemmerer W.A., Deiss S., Schwarz U.;		
RA "Cloning of Hyperion.";		
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.		
RN [5]		
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).		
RC TISSUE=Gastric parietal cells;		
RX MEDLINE=99115654; PubMed=99159845;		
RA Schmitt P.H., Dransfield D.T., Claudio J.O., Hawley R.G., Trotter K.W., Milgram S.L., Goldenring J.R.;		
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein associated with centrosomes.";		
RL J. Biol. Chem. 274:3055-3066(1999).		
RN [6]		
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).		
RC TISSUE=Lymphoblast;		
RA Hinds K., Sutterer C., Becker M., Hawkins M.;		
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
RN [7]		
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).		
RC TISSUE=Lung;		
RA Milgram S.L., Goldenring J.R., Schmitt P.H.;		

```

    )
    anchor
  )
  ized
  il;

  g M.;
  tat
  t

  choring

  Ono Y.;
  that

```


Db	330	----	NSNKEEIQE	-----	TIIEFLTNTKILIEEKKLELKD	365			
Qy	560	TTTS-----	IQPNLENSNSSSEINSS	OSSESAAAD	-----	PENGERSHTYVS 603			
Db	366	TADKLGLBLOQ	YQKNQJLKNKMLKELTNS	KQERQSSSEIQ	LNQNGVBLQKRHHKSO	425			
Qy	604	IQEVLVDFTSEKSTG	LSLSPGAGSGASTRT	ITRLRNPDSKLSQ	LKSQOVAANAHA	663			
Db	426	FETIVORME	-----	TORKLEPRL	-----	AEIDEMGOIV----- 458			
Qy	664	NKLREGEVLYVNS	QGEISRLSTKKEVIMKGINNY	FLQGEKRYVHNQYNSPAL		723			
Db	459	-----	QMKOELLIRHQAOMEKMT	RHREGENALRYSNI	-----	TYNEODIKLMAVI 507			
Qy	724	NKHORH-EDHDKRRH	-----	LAHKFCLTPAGEFKNWGS	VHSGKVLITLRLTITQ	773			
Db	508	NELMIKLTQDYN	SQKEKKEBLGLILBEKCALQ	-----		RQ 541			
Qy	774	LENNIPSSFHPNMA	SHRAMWIKAVOMCSKPRE	FALATALILECAV	PVYMLPIWREFLGH	833			
Db	542	LEDVEELSESREQ	ITORAROTI	-----	AQESKLNAHKSITVED	LAKAETISASESKLEL 599			
Qy	834	TRLHRMTS-----	TERREKEXKVKKEKQOE	EELMOATVKTTPYKQW	YMKOE	---E 886			
Db	600	KHEAEVTNYKIKLE	MEKENAVLDRASQOE	ELERLTQOLLSHEE	LSLKDLBLE	659			
Qy	887	YRVAGYGSWIS	KTTHVYRFPKLPQNTVNY	RKSLGKRNNDENMD	-----	ESDKRCSRS 944			
Db	660	HRIN-----	-----	IEKTKNLGHTHYQO	LDIGQONEMSKITELME	QERKNLITK 703			
Qy	945	PKKKTIEPDEK	DEVKGSAAKAD	-----	QNMDSIKTEKK	---DQVKEU- 989			
Db	704	QONLILETSKLD	QSLVNSKSEMTQ	INELQKEIEILRQ	EEKEKGLFQGEVQLQK	763			
Qy	990	-----	LDSDDKPCKEPRM	VDMDKTESHVNOQES	OYDVYVW 1028				
Db	764	TELLEKOMKE	ENDLOEKFALAE	-----	NSILKEKTKLEMLKIH	PPVSOE----- 813			
Qy	1029	SEGHLTSYK	KKTKSSKLDGLLE	RRIRKQPTLEEKORLEK	-----	IKLEGGJGIGIKTSSS 1086			
Db	814	-----	KLIFLIDTSKSKRS	SVWEKEL-ELIFPEN	DLKOQCIQINEL	EKORNTFSAE 866			
Qy	1087	KNL-----	SESPIVITKA	KEGQSDSMROE	-----	QSPNANDQPEDLIQGC	SOSDSSV 1134		
Db	867	KNEFVYQELQ	EYACLLKQDLEDSK	KNQBLEYKSK	KALNEE	-----	911		
Qy	1135	LMSDP	SHHTNKLYPKRVL	DVYSINS	PET-----	KCPKONS	TENDI 1176		
Db	912	LHIDRINPT	TVYKM-KSSVDE	DEKTEVAETLE	MGVEVEKDTLE	MKEKLEVT	REKLE--L 967		
Qy	1177	EEKVSDLAS	KGQEPTEKTKGN	FLIDDSKLMSAD	IGLILCKNNKPLIQ	ESDITVYSS 1236			
Db	968	SQRLSDIS	-----	DQKKHGEIS	FLNEEYKSLQOE	-----	KEOYSLRCLELITIHNN 1016		
Qy	1237	KSAHSSVPK	STNDROTPLSRAM	DEGKLGCDSES	NSTLENSSDTVS	IODSSEEDMIQ	1296		
Db	1017	RA-----	-----	ENVOCDTQ	YSLL-----	DGVYTMNS	SRGEGSVS 1048		
Qy	1297	NSNSIS	EOFPTEQDVE	LEPLKCELV	SGESTGNC	---DRLP--	VKGTEANGKRP	PSQ 1351	
Db	1049	KVNSFGE	ESKINWEDV	SPF-----	NMTYGE	ESKQOLILDLPL	PSVTKESSIR	ATOPSEN 1104	
Qy	1352	KKLEER	VVNCSDQIK	KANTTDKKN	NENESSEK	QGR-----	TSPFOI-----	NKQD 1398	
Db	1105	DKLO-----	-----	KELNLKSE	QNDLR-LQNEA	QRICLSLVSTH	DVQREY	WENKQD 1152	
Qy	1399	NKPRY	LKGCLKEIS	SRVSVS	GNVPKYNINIK	IIPENDIK	SLTVKESAI	RPINCQ- 1456	
Db	1153	-----	KALCSIK	EEILFA--QOE	KIKELQK-IHQ	ELODTMTQ	ET-----	GDG 1193	
Qy	1457	-----	VIMEDEN	ERNRSTSK	SHLLSSD	AGBNYRDS	LETLPTK	SDSTOTTPPS	ACPS 1512

Dd	1194	KPHLLIGLKOKAVSECCSFLOTLSVGEY-----	YTPALKC-----	1235	
Qy	1513	NSVNOVEDMEIETSEVAKVYTSPTITSEESNLNDEIDENGPIPKNN--	ENVNGESKRK-----	1565	
Dd	1233	-----	EVMNEDKEN -SGDIYISENDEPELODYRYEVOPOENMH	1265	
Qy	1570	TVITEVY-----	TMSTYATESKVIYKXEKDKOTVYSTENCAKSTVTTTTTV	1619	
Dd	1270	TLKNVTEEVYKLLVLOTRLSKITWGOOTDMKLEFGE-----	ENLPEK-----	1312	
Qy	1620	TKLSTPVGSGVDIISYKEDSKVYTVTTVDSLTITGGTLVSMYVSKYSTRDVKVLM		1679	
Dd	1313	-----ETPELSHSC-----	MTNLEIDIVNHSKSLSS	1335	
Qy	1680	FSRPEKTRSGTALPSYKRFYKTSKHSIFVLPNDLKLARKGIGREYVFNYPNAKPALD		1735	
Dd	1340	LQDLEKTKRLEQVOELLESLSLQOO-----	LKEHEQ-----NYEAF-----	1376	
Qy	1740	IMWYSPRPPTFGITWRYRLQTVKSLAGVSLMLRLMASLRMDMAKVPBGSGSTRTETS		1799	
Dd	1377	-----IHCLQ -KRLQAVS-----	ESTVPP-----SLP	1397	
Qy	1800	ETETITTEITKIRBDYGVYGRIFECYKRIKICPIGVPEYTEKEMPTPOGRKGLRBSALRPKR		1855	
Dd	1398	VDSVATITESDAQRTMYGS-----	CYAKKN -DGTEFSGEGVKEETINIVKLEKYOGE	1456	
Qy	1860	ETPKQGTGPIIETWVA -EEBELMEIETFAEYREYKSK -AQAVEQAQKRLLEQOETVIAT		1917	
Dd	1451	QLEEEVAKVIVYSMIAVAQOTELSRISGKEHTASSKQAHVQOQGHFNEMK-----	L	1509	
Qy	1918	STSPSTSTSTISISPAQVWAPVIGSGVTTGIMVLLTVKVSBAVTEQOKNHFOTAT		1977	
Dd	1506	SODQIGFQTEFETVAVKFEKFEKPLSKELGHEKKEILLS-----	NSDPHD-----	1549	
Qy	1978	WVKQOGNSGVYQOVKVLIISSGTGSOQTFTSQPTATVTTIRPNTSGSGGTTSMQ		2037	
Dd	1550	-----IPEKDCVLT-----	SEEMSCKTFTVROSITHIDEISVSMDSASQL	1592	
Qy	2038	VITPQPIRPMGTVTRTPLOOSTLGAKIITRPVWVPGAPQVWMTQIRGQPYSTAVSAPN		2097	
Dd	1593	MLNBEQLEDKROELVRYQEHQOATELRLRAHNRKQREHREDOEL--	QDEIKRLNQL	1649	
Qy	2098	TWVSTPGOKSLTSASTSNISQSSASOPPRPOGOVKLMAOLITOLITOGHGNGLTVVIO		2157	
Dd	1650	AORSSIDENLVNRSERHVLLELALKQJLSLAGREKLCCELNRNSTQJONGN-----	E	1702	
Qy	2158	GOGOTTCO-----	LQILPGV -TVLPFGGOOLMAAMPNGVQRFLLTPLATATATAS	2209	
Dd	1703	NQGEVEDQTEREKELDKRPDPPELISNERYALOKA--	NNRLKILLEVKTAAVEE	1759	
Qy	2210	T-----	TTTTVSTTAAGTGEOSRKSIPMOVHODKLP-----	PAOSSVGP----	2252
Dd	1760	TIGHVGLILDRSSKQSSASLIRSPAESVAKSCVHEEHTKVTDESIPISYSGSMPKND		1819	
Qy	2253	-----	AKAOPATOPASAPQOTOPASPAQE-----	VOTQPEVOTQTVSSHVPSEAQ	2301
Dd	1820	IMMSKVTEBGTELSQGLVNSFAGTEIPDENBELMLNISRLOAAVEKLLEAISETSSQ		1879	
Qy	2302	PTHAQSKPPOVAAS-----	QPOSNVQOGSPVAVQSPSOTRIRPSTPSOLSPGQ-----	2356	
Dd	1880	LEHAKVUTELMRESFROKQOATESLKCOBELERLHEBSRAREDLAVELSAEAGVIDGY		1939	
Qy	2351	-----OSOVOTTT-----	SOPRIPLOPHTSLQIPSOQOPQOSOPOVOSSTQTLSSGTL	2397	
Dd	1940	ADEKTLFERIOEKTDIIDLLEOELLCASIRLOELAE-----	QOOIOEBERLLSR-----	1990	
Qy	2398	NOVSVSSPRPOLQIOQPOPCVAVTAVPOLQOQOVVLSQIOSVVAATOAGSVPPOIK-L		2456	
Dd	1991	QKEMKMAEAP-----	VEQ-----	QLODETKMLKELVYQOQAEKVYRDLOKQYVAL	2038
Qy	2457	QLPITQIOSSAVOTHOIONV-----	VTVOAASVOEOLQRYVOLRDOQKKKQOOIEINVT	2512	
Dd	2039	EIVDEPOVSFEIEQKRNTELMDLRQONQALQOKEKRRKFLQADIRREHRRVGOE		2098	

DR PDB; 1MNE; 17-AUG-96.
DR PDB; 1VOM; 23-DEC-96.
DR PDB; 1VYK; 28-JAN-98.
DR DictyDb; DD01008; mhca.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
FT DOMAIN 762 791 IQ.
FT NP_BIND 817 2116 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 APP.
FT DOMAIN 638 660 ACTIN-BINDING.
FT DOMAIN 738 752 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678 ALKYLATION (SH-1).
FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MCK).
FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MCK).
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1EE56A1 C6C64;

Query Match 2.5%; Score 374.5; DB 1; Length 2116;
Best Local Similarity 19.1%; Pred. No. 2.3e-06;
Matches 356; Conservative 293; Mismatches 635; Indels 577; Gaps 81;

QY 31 DDDSDYEEMEDDDDDSYCTESSF--RSHSTYSTPGCRKPRHRRSPLE----- 81
DB 530 DEQSVFNP-----ATDNTLTIRKLHSHFS---KNAKYEPRFSKTEGVTNHYA 574
QY 82 -----EKDIPPLE-----FKSSEDLKVPNEHINVAIVEVLNFGTVLRLS 124
DB 575 GQVWYETIODEMEKKKDLQDLDELCKDSSDNVYTKLFNPENISRAKKGANFTTV---- 630
QY 125 PFRFEDCALVSOEQCTLAEMHVVLLKAVLREEDTSNTFG-----PADLKDS 174
DB 631 -----AAQYKQLASLMTL-----ETTNPHFVACIIPNKKQLPAKLEDK 670
QY 175 VNSTLYFDGTMPEVLR-----VYCESDKVHNHLYP--OEAE----- 212
DB 671 V-----VLDLRCNGVLEGRITRKGFPPNRITIVADYVKRYLLAPNVPDRAEDSQKATDA 725
QY 213 -----YRYGPVENKIKVLOFLVDQFLTNIAREELMSGV--IYDDHCRCYCHN 259
DB 726 VLKHLNDIDPEGRFG-----ITKIFRAGQGLARIEBARREGRISEIIKALQATGCGWARK 780
QY 260 LGDLLCCECTGSAVYHLBCVKRPLEEVPDEMQCEVCVAHKVPGVTDCAVLQCK-KPYIR 318
DB 781 -----VY-----KOAREHTVAARI-----IQQLRAYI- 803
QY 319 HEPIGYDRSRKKYWF--NRRLITEEDTENENKKIMYSTKQVLAELIDCDKDWEA 375
DB 804 -----DFKSPVWKLFKSKAPPLKRNFEK----- 829
QY 376 ELCKILEEMREELIRHNDITEDLTNKAAGSKSFLAANEELIESIRAKKGDIDNVKSPE 435
DB 830 -----IKKEKEELIELKSNLIDSTQK-----DKLEKSLMDTESNVLDLQRL 872
QY 436 ETEKDKMETENDSKD---AEKNREFEDQSLKSDDKTPDDDEQKSEPEPEVGDGN 492
DB 873 KAEKETLKAMDSDKDALEAQKRELEIRVEDMESELEDK-----K 911
QY 493 SVSANLQDNTNTAETSETSSEGRSPVGCSETPDSSNMAEKYASLPQDVPPEPKTK 552
DB 912 LALENL--ONOKRSVEEVRDLTEE-----LOEOKLKNLTLE-KLKRYKEELEEEMKRVND 964

QY 553 ESSNTSATTTSIOPNLNENSSSELNJOSESAAKADDPENGRESHTPVSIQEIYGVDF 612
DB 965 GQSDTIRLEKINDELO--KEVELTESPSESK-----DKVLEK--TRVRLQSEL--DDL 1015
QY 613 T--SEKSTGELSESFGAGSGASTRIITRLRNPDLSQLKSOQVAAAHAANKLFKEG 670
DB 1016 TVRLDSLTKDKE--LRRQKKLLEELKQVQELAAETAKLQEA 1059
QY 671 KEVLVNSQGEISRLSK--KEYIMKGNINNYEKLQGEKRYRVHNQYTNNSPALNHOH 728
DB 1060 ANKKL--QGEYTELNKEFNSEVTARSNV-----EKSRTLESQLVAVNNELEDEKKN 1109
QY 729 REDHDKRRHLAHKFCITPAGEFKWNG--SVHSGYL-----TISTRLITQOL 774
DB 1110 RDALEKKKALDAMLEEMKQOLESTGEEKSLVDIAKQKQSDMEALRNQISELQSTIAKL 1169
QY 775 ENNIPSSFHHPNMAHRANMIKAVQMSKPREPALLALILECAVPRVMLPIVREFLGHT 834
DB 1170 EK-----IKSP-----LEGEV-----A 1181
QY 835 RLHMTSIEBEKEKVKKKKKOE--EETMQOA--TWKKYTFPVKHQVKKQGEERYR 889
DB 1182 RLQGELEAEOLANSVNEKQKKVLEDEKSAQLAETAAKQALDKLKKLEOELSEVO- 1240
QY 890 TGYGQWMIKTHYRVPKLPQNTNVN---YRKSLEGTKNNDENMDESDEKRSKSPK 946
DB 1241 -----TQLEBAN-----NKNVNSDSTNKHLETSPNNKLLELAQKAKQKQALEK 1284
QY 947 KIKIIPSEKDEYKQSDAKGADQNE--MDISK--ITEKKDQVYKELLDSQPKCKEPP 1002
DB 1285 RLGLESELKLVNQEOLJEEEEKQKESNEKRYDLEKEVSELDQIEEBAKRAVTEAKN- 1342
QY 1003 MEVDDDMKTESHVNCQESSQVDVNVNVEGF--HLRTSYRK---KTVSSKLQGLER-- 1053
DB 1343 -----KKSELDDELKROYADAVVSSRDKSVOLKTLQAKNEELRNTAEABQDLRAER 1395
QY 1054 --RIKQPLF-----EKORLEKIKLEGIGKI-----GKTSTNSKNLSSES----- 1092
DB 1396 SKKAAEEDLEBAVKNLBEETAKVKAKEMAKKKAETDYSRTKSELDAKNVSSBOYQIKR 1455
QY 1093 -----PYITKAKGCOSDSMRQOSPANNNDQPEDLIQCSQSQSSVLAWS----- 1138
DB 1456 LNEBELSELRSVLEADRCNS-ALKARKTAESALESLEKDEIDANNAKAAEERSKVELEY 1514
QY 1139 -----DPSHTTK--LYPKDRVLDDV-----SIRSPETK--CPKQ----- 1169
DB 1515 RVALESELDEKSGTVVETIRKKDAEIDDLRARLDRETESRKISDEDKNTKQFADLE 1574
QY 1170 -----NSIENDIEEKVSDIASRQGEPTKSGTGNDFIDDSK-----LA-- 1208
DB 1575 AKVEAQREVVATIDRLKKLESIDIIDLTQLDIETKSRK-----TEKSKKLEQTLAER 1629
QY 1209 -----SADDLGTLICANKKPLIOE--ESDTIYSSKSALHSS-----V 1244
DB 1630 RAAEEGSSKADE-----BIRKQVQEWDELRAQDLSEFAALANASAKIKTSIQAIVEY 1683
QY 1245 PKSTNDR-----DATPLSRAMDE-----GKLQDSSENSSTLENSSD--TVSIQDSSEED 1292
DB 1684 KEQLEDEETIAKDKLVAKKRLAEVLEVRQOLEEEDSRBELDSKRLTLTYVEDIKKKY 1743
QY 1293 MIVQNSNESISQOFRTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTENGAKKSPQK 1352
DB 1744 DAVEQNTKLEDAKKKLTIDVDTLK-----KQLEDEKKKLINSE 1782
QY 1353 KLEERPYNKCSQDQIKLNTTDKKNENRESEKKQGRSTQINGKDNKPKITYLKGECLKE 1412
DB 1783 RAKKRLSENEDEPL-AKLDAEVKNNRSRAEKDR-----KYEKDLKDTKYK--LNDRAARK 1834
QY 1413 ISGSRVYSGNVBPYVNNINKIIPENDIKSLTVESAIRPINDVIMEDENENSSSTKS 1472
DB 1835 -TQTEIGAANKLEQIDELRSKLEBOAKATQADRS--KTLTBEIDNMLRAQIEDBGKIKM 1891

OY		1473	HLLSSDA-EENYDLSLELPKTRKESDTOTTPPSACSPBNVNOVEMELETSE--v	1522
Db		1892	RLEKERKLEEELEBELRETBEEADSKS-----EAFOSKRVLLEELEDARMLQ	19440
OY		1529	KKVSSPTTSEEENLNDNDIDENG-LPIKNENVNGESKRRTVTREVMTSTVAATESK	15870
Db		1941	KEIDAKELAEADKSNLORELVEAKGRLEESIKRTNSDKRKLAEIDALTAQYDAEK	20060
OY		1588	TIVKEGDKQTQVVSSTENCAC-----SVTTTTTVTLTSPPSTGSVDIISVEQSKT	16420
Db		2001	AKNQIKENKKIETELKEYKKKGEGSESKTKTEKFLIVEKLETFYKRAKKAENDQQQRUT	20660
OY		1643	V 1643	
Db		2061	V 2061	
<hr/>				
RESULT 12				
ATRX_MOUSE				
ID	ATRX_MOUSE	STANDARD;	PRT; 2476 AA.	
AC	061687;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Transcriptional regulator ATRX (X-linked nuclear protein)			
DE	(Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38 protein).			
GN	ATRX OR XNP OR HP1BP2.			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98213653; PubMed=9545503;			
RA	Picketts D.J., Tascian A.O., Higgs D.R., Gibbons R.J.;			
RT	"Comparison of the human and murine ATRX gene identifies highly			
RL	conserved, functionally important domains.";			
RM	Mamm. Genome 9:400-403(1998).			
RN	[2]			
RP	SEQUENCE OF 325-1176 FROM N.A.			
RX	MEDLINE=97133299; PubMed=8978696;			
RA	Le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,			
RA	Jeanmougin F., Losson R., Chambon P.;			
RA	Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,			
RA	Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;			
RT	"Localization of a putative transcriptional regulator (ATRX) at			
RT	pericentromeric heterochromatin and the short arms of acrocentric			
RL	chromosomes.";			
RM	Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).			
CC	-1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES			
CC	GENE EXPRESSION BY AFFECTING CHROMATIN.			
CC	-1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXXIN V IN A CALCIUM AND			
CC	PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By			
CC	similarity).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC			
CC	HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY			
CC	INTERACTING WITH HP1.			
CC	-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstation at			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			

[illegible]

```

QY 459 -----EDOSLEKSDDKTPDDDPQSGKSEPT-----EVDGKNSVSA 496
D 459 LKIDARSSVKAIDGEEBRAHSTGHEKSGRGDSOYEPTNTPEDLDMDIVSPSSVPE 518
QY 497 NLGDNNTNATSEETSPS-EGRSVGLSETPDS--NMAEK-----KVASEL 540
D 519 DIDPSLEAMEVQSSADYQDNGSTPELESSVYKLVNSKDSRGNIKSKYAKAVKEL 578
QY 541 -----PODVPEEPNKTCESSNTSATTTSTIOPNLNSSSEL----- 577
D 579 FVXLTPVSLNSPIKGYVCCOEVSQEKNGR-KSSGVARSSEKCRPREISDHENNVITLLE 637
QY 578 NSSQSEAKAADDPENGERSHRPVSTQIEIYQDPTSEKSTGELSESPGAKGASGSTR 637
D 638 DSDLRSPRYKTPPLRQTSNPNAMSND-----EESNGTKER----- 676
QY 638 IYRLRNPDSKLSOLSKSOVAAAHAANKLEKGEKVLVNSQGEISLSTKKEVIMKNI 697
D 677 -QKMSGPIRKDKRNSADCATDNPKPKHVPKAKQPIGDNDSDEMLAVLKEASQMGHS 735
QY 698 N-----NYRKLQGECKYRYHHNYSTNSFALNKHQREDHDKRHLAHFCLPAGEFK 751
D 736 SSSDTDINEPQMHGK-----TGKDNQGR-----K 762
QY 752 WNGSVHGS-----KVLITSLRLITOLENNIPSSFPHPWASHRAMWIKAVQMSKPRE 806
D 763 RKNSTSGSDPTKKGKSTETSIISKRRQVSESS-----NYDSELEHETMSRIGAKR 818
QY 807 FALALAILCAVFPVMLPIWREFLGHTRHMTSIREKEKVKKKEKKO----- 857
D 819 -----SVP-----EKKEEDSSDEKOGKKVYDNGG 843
QY 858 -EEBETMOATWVYTFPVKHQVWKQGEERYVTVGYGWMISKTHYRVPKLPG----- 912
D 844 HERAKTTQEGSSADDTGDT-----GROGSGCSLAG-GSIEKVSQGVFRMLCKPQVSSD 898
QY 913 -----NTNVN-----YRK-----SLEGT-----KNNMDENMD 934
D 899 GAEPKSVKEENVSPEDKRVSKTEKTKHLRSQRKRGKGGSDGTDRFPKQESDESSE 958
QY 935 ESDKRKCSRSRKKIKIIPDSKDEV-----GSDA-----AKGDQVEMD 974
D 959 GEKKQSRQRPQTKKAPDLGELTKREQWESSSDGTERLPEELETGIPSPKIGQSKTD 1018
QY 975 IS-----KITEKKDDVLELDDSDKPKCEPMVDDMKESHVNCQESSQVAVN 1027
D 1019 TAGGEKKGKKWKKSCCKEKKELSDVYKLPKGDCSSSEDKTRNNVSLREKKRFSLLPA 1078
QY 1028 VSEGFHLRTSYKKRTKSSK---LDGLERIKOPTLEKORLEKIKLEGIGIKGTSTN 1084
D 1079 KSPKRECCSSSDTEKSLKGCSCDS-TEKRPKRIDLERR-----NSSSKRNTK 1126
QY 1085 SSKLSESPIVITKAKBEGQSDSMROBOSPANNNOQPDLLQGCQSDSSVLRMSDPGHTT 1144
D 1127 EYKSAHSS-----SPAEGSSEDNKQKKO-----RTAKKKTG 1159
QY 1145 NKLYPKRVLDDVSIRESPTKCPQNSIENDIEEKVSDLASRQGEPTKSKTKGNDPFIID 1204
D 1160 N-----TEKKRNSLRAPPKRQVDITS-----SSSDIGDD 1191
QY 1205 SKLA-----SADIGILCKKNKP-----LIQESDTIYSSKSAHSSVPKSTNRD----- 1232
D 1192 QNSAGEESDSE-----QIKIPVTENIYLPSTHGFCCOSSGGEALSKSPAVAVDDDDND 1245
QY 1253 -ATPLSRAM--DPEKRGCSSESNTLENSDVTVSIDQSEEMIVONSVESTSEQPT 1308
D 1246 PENRIAKMLIEITKANSSDEGSSDDE-----PDGGGKRIRIGQSESPADGCEL 1237
QY 1309 REODVEVLEPLKCLVSEGSTGNCEDRLPYVGTANGKKPKSQOKL-----E 1355
D 1298 RREOLAVNQ-----VNSSEDSDEE-----SKPRYRRLRLRHKLTLTLDGESGE 1341

```

```

QY 1356 EREPVNKCSDQIKLKNNTDKNNENRESEKKGQRTSTFOINGKDNKPKIYLKGBCLKEISE 1415
D 1342 EKPT-----KPKHREKAKGRNR-----KVSDESDTDQESGVSEVSE 1382
QY 1416 SRVYSGVNEPKVNNINKIIPENDIKSLYKESAIRPPIGVDVIMEDFERNSETSHLL 1475
D 1383 S-----EDEORPRTSARKAELEENORSYKOKKKRRRRTKVOED-----SSSENKSH-- 1428
QY 1476 SSSD-ABGNRYD-----SLETLPTKESDSYQTTTPASCPESNS 1514
D 1429 SEEDKKEGDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1488
QY 1515 VNQVEDMELETSEVKRY-----TSSPTSEESNLSDNDFIDENGDLINKNENNGES 1566
D 1489 RKRIAEERERERKLEVEITIEDASPTKCPITTK-----LVLEDEE----- 1529
QY 1567 KRRTVITEVTMTSTYATEKTYIKYKDGDKQVVSSTENCAKSTVTTTTTYTKLSTPS 1626
D 1530 -----TKBPLVOVHRNVIKLPHQVDGVOEFMDDCCES-----VEKTKRSPGS 1573
QY 1627 -----TGGSVDIISVKEQSKTVTTVTDSLTGTGTLV-----TSMTVSKY----- 1669
D 1574 GCLIAHMGIGKTLQVVSF-----LHTVLQDKLPSTALVYCPILTALNMNNEFEKNO 1627
QY 1670 ---STRDKVKLMFSRPKKTRSGTALPSTRKVEYTKSKSIFVLPMDLKLAKKRGIRE 1726
D 1628 EGLNDEKLEVSBLATVVR-----POERSYM-----LQRMQEDGVMI 1665
QY 1727 VPFYFN 1732
D 1666 IGYEMY 1671

RESULT 13
PHP_DROME STANDARD: PRT: 1589 AA.
AC P39769;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polyhomeotic (proximal) chromatin protein.
GN PH-P.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Braachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE=imaginal disks;
RX MDLINE=92146957; PubMed=1346609;
RA Decanilis M., Cheng N.S., Pierre D., Brock H.W.;
RT "The polyhomeotic gene of drosophila encodes a chromatin protein that
RT shares polyene chromosome-binding sites with Polcomb.";
RL Genes Dev. 6:223-232(1992).
RN [2].
RP SEQUENCE OF 199-1584 FROM N.A.
RX MDLINE=92039031; PubMed=1937015;
RA Beatrix J., Daly M., Randsholt N.B., Brock H.W.;
RT "The complex genetic locus polyhomeotic in Drosophila melanogaster
RT potentially encodes two homologous zinc-finger proteins.";
RL Gene 105:185-195(1991).
CC - FUNCTION: BINDS TO POLYENE CHROMOSOMES. SEEMS TO INTERACT WITH
CC PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER
CC COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND
CC SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF
CC OTHER PAIR-RULE GENES SUCH AS EVE, FTY, AND H.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: SALIVARY GLANDS.
CC - SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC - SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MEN-9 IS THE INITIATOR.

```


[3]
RP SEQUENCE OF 1-1909 FROM N.A.
RX MEDLINE=93390935; PubMed=8378076;
RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
Kodera Y., Nakawara S., Ariyoshi Y., Takahashi T., Ueda R.;
RT "Two distinct portions of TRG19/ENL at 19p13 are involved in t(11;19)
leukemia.";
RL Oncogene 8:2617-2625(1993).
RN [4]
RP SEQUENCE OF 1317-2328 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93265134; PubMed=1303259;
RA Diabali M., Sellert L., Parry P., Bower M., Young B.D., Evans G.A.;
RT "A trithorax-like gene is interrupted by chromosome 11q23
translocations in acute leukaemias.";
RL Nat. Genet. 2:113-118(1992).
RN [5]
RP SEQUENCE OF 1251-1538 FROM N.A.
RX MEDLINE=94215165; PubMed=8162575;
RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Cnaan O.,
Saito H., Croce C.M., Cnaan E.;
RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
involved in acute leukemia.";
RL Cancer Res. 54:2326-2330(1994).
RN [6]
RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX MEDLINE=95322025; PubMed=7598802;
RA Mongkolkeo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
Rowley J.D., Diaz M.O.;
RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
trx zinc-finger domain, and alternative splicing.";
RL DNA Cell Biol. 14:475-483(1995).
RN [7]
RP SEQUENCE OF 1212-1603 FROM N.A.
RX MEDLINE=95315013; PubMed=7794749;
RA Marschalek R., Grell J., Lochner K., Nilson I., Siegler G.,
Zweckbronner I., Beck J.D., Fey G.H.;
RT "Molecular analysis of the chromosomal breakpoint and fusion
transcripts in the acute lymphoblastic SEM cell line with chromosomal
translocation t(4;11).";
RL Br. J. Haematol. 90:308-320(1995).
RN [8]
RP SEQUENCE OF 1421-1540 FROM N.A.
RX MEDLINE=94020842; PubMed=841518;
RA Forster A., Rabdlits T.H.;
RT "A method for identifying genes within yeast artificial chromosomes:
application to isolation of MLL fusion cDNAs from acute leukaemia
translocations.";
RL Oncogene 8:3157-3160(1993).
RN [9]
RP CHROMOSOMAL TRANSLOCATION WITH GAST7.
RX MEDLINE=20183971; PubMed=10706619;
RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
Williams T.M., Lange B.J., Felix C.A.;
RT "Detection of leukemia-associated MLL-GAST7 translocation early during
chemotherapy with DNA topoisomerase II inhibitors.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
RN [10]
RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
RX MEDLINE=20115194; PubMed=10648423;
RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
RT "Novel SH3 protein encoded by the AF3P21 gene is fused to the mixed
lineage leukemia protein in a therapy-related leukemia with
t(3;11)(p21;q23).";
RL Blood 95:1066-1068(2000).
RN [11]
RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL,
T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
CC THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES

CC MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
CC MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;
CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;
CC T(11;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)
CC that involves MLL and ELN; t(11;19)(q23;p23) that involves MLL
CC and GAST7; and t(3;11)(p21;q23) that involves MLL and AF3p21.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
CC DATABASE: NAME=Atlas Genet. Cyto genet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MLL.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04284; AA58669.1; -;
CC EMBL; Z69744; CAA93625.1; JOINED.
CC EMBL; Z69745; CAA93625.1; JOINED.
CC EMBL; Z69746; CAA93625.1; JOINED.
CC EMBL; Z69747; CAA93625.1; JOINED.
CC EMBL; Z69748; CAA93625.1; JOINED.
CC EMBL; Z69749; CAA93625.1; JOINED.
CC EMBL; Z69750; CAA93625.1; JOINED.
CC EMBL; Z69751; CAA93625.1; JOINED.
CC EMBL; Z69752; CAA93625.1; JOINED.
CC EMBL; Z69753; CAA93625.1; JOINED.
CC EMBL; Z69754; CAA93625.1; JOINED.
CC EMBL; Z69755; CAA93625.1; JOINED.
CC EMBL; Z69756; CAA93625.1; JOINED.
CC EMBL; Z69757; CAA93625.1; JOINED.
CC EMBL; Z69758; CAA93625.1; JOINED.
CC EMBL; Z69759; CAA93625.1; JOINED.
CC EMBL; Z69760; CAA93625.1; JOINED.
CC EMBL; Z69761; CAA93625.1; JOINED.
CC EMBL; Z69762; CAA93625.1; JOINED.
CC EMBL; Z69763; CAA93625.1; JOINED.
CC EMBL; Z69764; CAA93625.1; JOINED.
CC EMBL; Z69765; CAA93625.1; JOINED.
CC EMBL; Z69766; CAA93625.1; JOINED.
CC EMBL; Z69767; CAA93625.1; JOINED.
CC EMBL; Z69768; CAA93625.1; JOINED.
CC EMBL; Z69769; CAA93625.1; JOINED.
CC EMBL; Z69770; CAA93625.1; JOINED.
CC EMBL; Z69772; CAA93625.1; JOINED.
CC EMBL; Z69773; CAA93625.1; JOINED.
CC EMBL; Z69774; CAA93625.1; JOINED.
CC EMBL; Z69775; CAA93625.1; JOINED.
CC EMBL; Z69776; CAA93625.1; JOINED.
CC EMBL; Z69777; CAA93625.1; JOINED.
CC EMBL; Z69778; CAA93625.1; JOINED.
CC EMBL; Z69779; CAA93625.1; JOINED.
CC EMBL; Z69780; CAA93625.1; JOINED.
CC EMBL; D14540; BAA03407.1; -;
CC EMBL; L01986; AAA92511.1; -;
CC EMBL; U04737; AAA18644.1; -;
CC EMBL; S78570; AAB34770.1; -;
CC EMBL; X83604; CAA58584.1; -;
CC EMBL; S66432; AAB28545.1; -;
CC EMBL; AF231998; AAG26532.2; ALT_TERM.
CC TRANSFAC; T02337; -;
CC Genew; HGNC:7132; MLL.
CC MIM; 159555; -;
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR003889; FYTich_C.
CC InterPro; IPR003888; FYTich_N.
CC InterPro; IPR003616; Postset.

DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; znf_CXXC.
DR InterPro: IPR001965; znf_PHD.
DR Pfam: PF00628; PHD. 3.
DR Pfam: PF00856; SET. 1.
DR Pfam: PF02008; zf-CXXC; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD. 2.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; zf_PHD_1; 3.
DR PROSITE: PS50016; zf_PHD_2; 3.
KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;
Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
KW Alternative splicing; Repeat.
FT DOMAIN 17 102 ALA/GLY/SER-RICH.
FT DNA_BIND 169 180 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 217 227 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 301 309 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1147 1194 CXXC-TYPE.
FT ZN_FING 1431 1482 PHD-TYPE 1.
FT ZN_FING 1479 1533 PHD-TYPE 2.
FT ZN_FING 1566 1627 PHD-TYPE 3.
FT DOMAIN 1703 1748 BROMODOMAIN (DIVERGENT).

Query Match 2.5%; Score 367; DB 1; Length 3969;
Best local similarity 17.6%; Pred. No. 8.7e-06;
Matches 620; Conservative 405; Mismatches 1211; Indels 1278; Gaps 153;

QY 5 EEEEDGDAEETODSDDEDEDEDDDDSDYEEMEDDDDDSYCTESS----- 54
DB 856 EELSKRDGDAEVEKSKRERERERENK--RESREKRRKSGEIQSSSALLYPVGRVSK 913
QY 55 -----FRSHSTYSSNPGRRKPRVH----- 73
DB 914 EKVGEDVATSSAKKATGKKSSSHDSGTDITSVLGDTAVKTKILIKKGNLEKTN 973
QY 74 ----RPSPIIEEDDIPPLEPKSSEDLMPNHEIMINVAIYEVLNFGVLRISPEREED 130
DB 974 LDGPAFSPLEKEKTEKLTCLSTPSSS-----TVKHSTSSIGSM----- 1010
QY 131 FCAALVSOECTLMAEMHYVL-----KAVLEEDTSNTTF- 166
DB 1011 -----AQADKLPMDKRVASLTKKAKAOLCKTEKSKLKQTDQPKAQQGSSDSETSVR 1064
QY 167 GP-----ADLKSYNSTLYFTDGMTWPE--VLRYCESDKEYHHVLP 206
DB 1065 GPRIKHVCRRAVALGRKRAVFPDDMPT-----LSALPWBEREKILSMGMDSS--SIAG 1118
QY 207 YQAEADYPRGPEVNAIKVLOFLVDQFLTNIAEELMSEGVIOYDDHCRVCHKLGLDLC 266
DB 1119 SEDAE--PLAPRPKPIK-----PVTRNKKAPQEPVPKGRSRRCQ-----C 1158
QY 267 EPGSAVYHLECVAPLEVEPEDEMOCEVCVAHKVPGVTDVAIQNKPKYIRHEPIGYDR 326
DB 1159 PGC-----QVPED--CGVC-----TNCL-----DKP----- 1177
QY 327 SRKKYFLNRRLIIEEDTENENKIMYVSTKVQLAELIICLDKDWAEALCKILEEMRE 386
DB 1178 ---KFGGRN-----IKKQCKMKKCKNLOMPS 1202
QY 387 EIHRRHMDITEDLTNKAAGSKSFLAANBEILSIRAKKGDIDNVKSPFEETDKMETEN 446
DB 1203 KAV-----LQKQAKAVK-----KEKSKTSEKKDSKSSVYKKNVDS 1240
QY 447 DSKDAEKNEEFEDSDDKTPDDPEQKSEPEPTVGKGSVSNLGDNTTAT 506
DB 1241 SQKPTPSARD-----DAPKKSSEPPPRKPFVEKSEB-----GN-VSAP-GPESQAT 1287

QY 507 --SEETSPSEGRBPVGLSETPDSSNNAEKKVASSELPQDVPDEPNKTCSSNTSATTTSI 564
DB 1288 TPASRKSCKQVSPALVIPPOPTTPPRK-----EVKTPPSEPKK----- 1330
QY 565 QPULNENSSSELNSSSEAKAADP-----ENGRESHTVYSIOELIVGDTSEKST 618
DB 1331 QP-----PPESGPEGSCKKVAAPRSIPVKQPKPEKRPVKNQEN----- 1373
QY 619 GELSESPGAKGASGRITRLRNPSKLSQLKSOQVAANAABANKLPFEGKEVLVNS 678
DB 1374 -----AGTINILSTLSNGNS-----KQIPADGVIRIRVDFEODEAEVWE 1416
QY 679 QGEISRLS-----TKKEVIMKGINNYFKLGQEKYRVYHNOYSTNPFALNKHOREDHK 734
DB 1417 MGGIGLITVPITPRVVC-----FLCASSGHVEVYQVOCCEP----- 1455
QY 735 RRIHAKFCITPAGERFKWNGSVHGSKVLTITSLRLITTOLENNIPSEFFPNASHRAW 794
DB 1456 ----HKFCLD-----ENERPLEDQLENNCCRCR- 1479
QY 795 IKAVQMSKRRERFALALALEC-----AVKPVVMLPIWREFLGHTPLR 838
DB 1480 -KFCVCG--RQHOATKQLECKNCKNSYHPECLGPNYPKPKKKKWTI-----CTKCVR 1532
QY 839 MTSIEREEKKVKKKKKQEEETMOQATWVXYTFPVKHQVMK--QKGE----- 885
DB 1533 CKSGGSTTPK-----GMDAQW-SHDFSLCHCAALFAKGNCPICDKCYDD 1578
QY 886 ---EYRYTGYGWS-WIS-----KTHVYRFVRLPGN--TVNVRK----- 920
DB 1579 DYESKMQQCGKCDRWVHSGCENLSDMEYELISLNPESAVYTCVNCTERHAPAMRLALEK 1638
QY 921 ----SLEGTNNMDENDDESDKRCSPKIKIEPSE-----KQEVKSD-----AA 965
DB 1639 ELQISLQVLTALNLSSTTSHLLRYROAKPRDLNPEFTESIPSRSPGCDPVLTEVS 1698
QY 966 KGADQNMEDISKITEKKQDQ-----VKELDSDSDPKCEPEPVEDDMK 1010
DB 1699 KODQDQPLDEGVKRMKDQGNITSYLEFSDDIYKIIQAALNSDGGQ-----EIKK 1749
QY 1011 TESHVACQESSQYDVYVNSGCHLRTS-YAKTKSKSLDGLERRIKQPTL----- 1060
DB 1750 ANSMVKSFFIRQWE--RVFVFWGVKSRFEPKPKVSSNSGMLNPAVLPSLIDNYAQOMD 1807
QY 1061 -EEKQRLKIKLGGI-----KGIGKTSNNSKNLSESPVIRKAKGCGQSDMRQOSP 1113
DB 1808 REENSHTEQPLPKKIIIPAKPKGPEPDSPTPLHPPTPIIL-----STRSRKDSR 1859
QY 1114 NAN-----NDQPEDLIQ-----CSQSDSSVLRMSD----- 1139
DB 1860 ELNPPGIEDNRQALCLITYGDDSANAGRLVYIGQNEWTHVYCALMSAVFEDDGSLLK 1919
QY 1140 -----PSHTTNKLYPKDRLVDVYSINSPEPKCRKONS 1171
DB 1920 NVHMAVIRGKOLRCERCQKPGATVGCITSCSNYHMCSTRANCVFLDKKYYCQORHRD 1979
QY 1172 I-----ENDIE-----EKVSDIAS--RGQETKSKTGNGNFIIDD-----SKIAS 1209
DB 1980 LKGEVVPENGFEVFRVAVDFEGISLRKFKLNGLEPENIHMMIGSWTIDCLIDLNDLSD 2039
QY 1210 ADD-----IG-----TLICK--NKKPLIQ-----EESDITVSSS 1236
DB 2040 CEDKLEPIGQCSRXYWSTTDARKRCVYICKIYECRPVYVEPPIINSTVEHDKRTIAHSP 2099
QY 1237 KSAHSSVPRKSTN-----DR-----DATPLSRANDF--EKKLGC 1268
DB 2100 TSFTSSSKESQNTAETIISPSPDRPHSQTSGSCYUHVYISKVRIPTPSYSPFORSPOC 2159
QY 1269 D--SENSSTLENSDVIYSDSSSEDMIVONSNEGSIQORFTEQDVEYLER--LACELY 1324
DB 2160 RPLPSAGSPPTTHETIVGDP-----LSSGLRSIG--SRHSTSSLSIPQSRKLRTI 2210
QY 1325 SGESTNCEDRLPVKGTGANGKKRPSQK--KLERPVKNKCSQDIKILKNTTDDKKN-- 1377

```

Db 2211 SPARTGTYSNNVSVSTGTATDLESsAKVVDHVLGPLNLSSTLGQNTSTSSNLTORTV 2270
QY 1378 ---ENRSEKKGQRTSTF-QINCKDNPK--IYLKGECLKEISEERAVSNGNEPKVNN--- 1429
Db 2271 VTGNKNSHLDGSSSEKOSASDLVSKSSSLGKTKVLS-----SKSEGAHNAVAY 2325
QY 1430 --INKIPE-NDIKSLTVKESAIRPINDGVIEMEFERN-SPTKSHLLSSDAEGNRY 1485
Db 2226 PGIRKLAPOVHNTSRELNYSKISFAPESSV--SFSSKELSPHLHLRQNRDRQHT 2383
QY 1486 DSLETLDPSTESDS-TOTTPSASCPESSVNO-----VEDMEIETSE 1527
Db 2384 DSTOSANSPDEDEVTYTLKLSGMSNRSIINEHMGSSSRDRROKROKSKCEPTEKHSS 2443
QY 1528 VKRVTSSPTSEESNLSNPIDE-----NGPLIKNE 1560
Db 2444 KSFLPEGOVTTGEGCNLKPEPMDEVLPETWGORPCNNVSDKIGDKGLSPGVKAPPM 2503
QY 1561 NVNGESK-----RKVTTEVTMTSTVATESKTIVIKVEKGKQTVVSTENCASVTYT 1614
Db 2504 QVESAKELQAPRKRYKVTLLPLKMESESQKNALK-----ESSPASPIQISTSS 2554
QY 1615 TTTTIVTLSTPSTGSGSDIISVKQSKTVYTTVTDLSITTTGGTLVSMYVSKETSTRDK 1674
Db 2555 PTEPISASENPGDG--PVAOPSPN--NTSCODSOSNNYONL-----PVQDR 2596
QY 1675 VKLKFRPRKPTRSGTALPSYRKFVTKSKSIFVLPNDLKPL--ARKGIREVPYF-- 1730
Db 2597 -NMLPQPKRQEDGSKFRKRPKRSARASRMF-----GLTPLYGRSAGEEDIPYSS 2650
QY 1731 -----NYNAKPAIDIMPYSPRPFTGITWRYRLQTVKSLAGVSLML 1771
Db 2651 STGKRKRSAGEQVDDADLSTDEDDLYYNNFR-----TVISSGEE-- 2695
QY 1772 RLMLASLRMDMAKVP-----GGGSTRTESETEITTELTILIKRDVGPICIREYCIRK 1827
Db 2696 RLASHNLFREEQCDLPKISOLDGVDGTE--SDTSVTAT----- 2733
QY 1828 IICIGVETPKETPTPOKRLIRSSA--LRPKREPTEKOTGCVLIETWVAEELEMEIR 1885
Db 2734 -----TKSSQIRKRNCKENGTEMLKIDREPDAKEHEHYKSSVGAKNEFKMCH 2784
QY 1886 AFAERVEKRAQAVEQAKRLEQOKPVIATSTTSSTSTISPAOKVMAVAPISGV 1945
Db 2785 SVSR-----YKTGQDSLEAQSSL-----ESSRVHSTSPSKNLD-- 2822
QY 1946 TTGKMYLTKVGS-----PATVFOQKNFHOTFATVWQGSNS-----G 1987
Db 2823 TYNTLELKSDNNNSDDCGNLLPSDMDVFLKNTPSMOALGESPESSSELNLGEGLG 2882
QY 1988 VVOYQOVVLGIIPSTGTSOOTFTSFQPRATVTRPNTSGSGGTTNSQVTTGQIRPG 2047
Db 2883 LDSNREKDMGF-----EVEFQQLPTEPRV-----DSSVSSISAEQFELPLELPSD 2930
QY 2048 MTVIRTPLOOSTLKALIRPVMVOPGAPOQVMTQIRGQPVSTAVSAPNTVSTPGQKS 2107
Db 2931 LSVLTJ-----RSPTV-----PSQNPBSRL-----AVLSDSEKR 2959
QY 2108 LTSATSNIOSSASOP-----PRPO-----QGOV-----KITMAQLIQLTQGHG 2147
Db 2960 VT--ITKSVASSSDPALLSPGVDPPEGHMPPDHFIOGHMDADHISPPCGSVEGCHG 3017
QY 2148 GNOGIIVVIOGQGTQGLIIPQVTV-----LPQGO----- 2181
Db 3018 NNQDLT-----RNSSTGLO--VPSPVTPVIONQKYVNSNDSPEPSISNAAVOTTPPHL 3071
QY 2182 -----QLMQ-----AAMPNGTVQRFLEPTPLATATTAATTATTTVSTTAAGTEQ 2225
Db 3072 KPATEKLIIVNONNQPIYVLOTLPNGVTKIQLTSSSVSPSVETMTSVLGPBG--GLT 3130
QY 2226 RQSLSLPOMOVHODKTLPPAOSSSVGPAK-----AOPQTAQPSARPOPOTOP----- 2272

```

```

Db 3131 LTTGLNLSPTSQ--SLFSPASKGLBPMHHHLSPAPTAQSSFPNNISNPPSGLLIGV 3188
QY 2273 QSPAQPEV-----QOPPEVOTOTVSSHVS--EAP--THAOSKPYVAASOPQSNVOG 2325
Db 3189 QPPDPOLVSESSQKRDLSITYAT--PSSGKLKRPISRLOTKRNKLAASSIPSNT--- 3243
QY 2326 SPVRVOS-----PSQTRIPSTPSQLSPG-----OOSQVQTTTSQLIP 2363
Db 3244 APSDVSNMMLINFTPSQL---RPHPSLLDGLSINTSHRTVPNIIRKSKSIMEFEPAP 3300
QY 2364 IQPH-----TSIQIPSOQPOSOPOVOSTOTLSSGOTLNOVSVSPRPOLQIQP- 2415
Db 3301 LLPQSVGTAATAAGTSTISQDTSHLTSGSVSGLAASSVLANVSMQTTTTPTSSASVPG 3360
QY 2416 -----OPVIAVPOLOOVVLSQIOSOVAQOAGSQPOQIKIQLPIQIOQSSA--- 2467
Db 3361 HVTLTNRLILGTPIGS-----ISNLIKASQSLGIQDO-----PALPSSGMFP 3407
QY 2468 -VOTHOIONVTVQASVQDQRLQVQQLRDQOKKQOQIEINVPNSKLLIKVEITIQK 2526
Db 3408 QLGTSQTPSTAATRAASSICVL-----PSTQTTGITRASPSS----- 3443
QY 2527 VVKHNNAVIEHLKOKKSMTAERENORMIYCNVMYIIDKIDKEKQAKKRRRESV 2586
Db 3444 -----GEADEHYQLOHVNOL----- 3459
QY 2587 EOKRKONATKLSALFLFKHKEQLRAETILKRALLDKDQIEVQELERDKIKKEDLMQ 2646
Db 3460 -----ASKTGT-----HSSQ-----RLDASASGQVSNFTQYDAPRISMG 3494
QY 2647 LAQATAVAAPCPVTPVLPAPRAPPSPPPPGVQHTGLST--PLPVASQKKRKEEK 2704
Db 3495 LEQNKALS-----SAQASPTSPGSGSPSSSQRSASPSVGPPTPKPKTRFRQLPLD 3548
QY 2705 DSSSKSKKKKMISTSKET-----KKDKIKYCICQTPVDESKFTGCDRCQNMWNGRCVGI 2760
Db 3549 KGNCKRHKRVSHLRTSSSEAHIPDQETTSLTSGTGPAAEA----- 3589
QY 2761 LQSEAEILIDEVYCPQOCOSTEDAMTVL-----TPLEKYEGLKRVLSRLQAHKMAFP 2813
Db 3590 -QDPTASVEDSSQKECGQAPQAVALEPVQYQNPABEQ----- 3527
QY 2814 LEPVDPADPDYGVIEKPMDLATMEERVQRRYEKLTFFVADMTKIFDNCRIYNPSDSP 2873
Db 3528 -ESAEPKTEVEEESNFSSPLMLWLOQOKRR---ESTTEKKPKKGLVFE-----ISSDDG 3678
QY 2874 FYQCAEVLSEFFVQKLGFKASRSHNNKLOSTAS 2907
Db 3679 FOICAESIED-----AMWSLTDKVOEARS 3702

```

```

RESULT 15
NCRL_MOUSE STANDARD; PRT; 2453 AA.
ID NCRL_MOUSE
AC 060974; 060812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Nuclear receptor co-repressor 1 (N-COR1) (N-COR) (Retinoid X receptor
interacting protein 13) (RIP13).
GN NCOR1 OR RXRIP13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10990;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE=Plutitary;
RA MEDLINE=96008539; Pubmed=7566114;
RA Hoerlein A.J., Naeaeer A.M., Heinzl T., Torchia J., Gloss B.,
RA Kurokawa R., Ryan A., Kamei Y., Soederstroem M., Glass C.K.,
RA Rosenfield M.G.;

```

*Ligand-independent repression by the thyroid hormone receptor

QY 1194 KTKNDEFIDSKLASADIGTLICKNKRPIL---OESEPTIVSSKSLALHSSVPKSTND 1250
Db 1111 -----APLTIKOEFS-----PRQNS 1128
QY 1251 RDATPLSRAMD---FEGKLCDSKSNSTLNSSDPTVSIODSSEDMIVQNSNESISBQFR 1307
Db 1129 QPEBLVRAQHEGVVGRGTAGAVQGSILTRGTPAKIS----- 1165
QY 1308 TREDDVELEPLKCELVSGETSGCEDRLPYKG--TEANGKKPSQOKKLEERPYNKCSQ 1365
Db 1166 -----VETISSLRGSIQOGTPA-----LPQAGIPTALVKGVPVSRNPLESSPEKREE 1214
QY 1366 IKLKNTP--DKKN-----ENRESEKKGORT--STFOINGKDNPKI---YLKCECLKEI 1413
Db 1215 AASGHVYIEBKSGHLLSYDNIKARREGTSRPTAHMSLKRSEAVEGSIKQMSKRES 1274
QY 1414 SESHVSGNVEPKVNNINKIIPENDIKSLVYKESAIRPFIINGDIYIM-----EDFNE- 1464
Db 1275 PVSAPLEGLI---CRALPRGSPHSDLKERTY-----LSGSIQGTPRATASPFEDGL 1323
QY 1465 -----RNSSETSHLSSDAGNRRDILETLPSTKESSTOTQTPPSACSPESNVSQ 1517
Db 1324 KYPROIKRESPIRA--FEGALITKGKPYDGIYTI--KEM-----GRSIEH 1364
QY 1518 VEDMEIETSEYKVV-----TSSPTSEESNLSDFIDENGLPIKNKNENVGESKRTVI 1572
Db 1365 IPRDILTOESKRTPEVQOSTRPII---EGSIS-----QCTPI-KPDNNSGQSAIKHNV 1414
QY 1573 TEVTMTSTVATESKTYVIK-----VEKQDKO-----TVVSTENCACKTV-- 1612
Db 1415 KSLITGPKSLPRGMLIEIVPENIKYVERGKYEDVAKAGEVBARHTSVVSSGSPVLRSTLHE 1474
QY 1613 -----TTTTVTYKLSPTSPSGSV-----DIISVKEQS--KTVVTTITVD 1650
Db 1475 APKQOLSGLYDDSSARKTPVSYOITISRGSPMANRTSDVSSSKASHKSTLTPTORE 1534
QY 1651 SLTTTG-----GLIYV-----SMTVSKYSTROKVKL--MKFSRPKRTSGTAL- 1692
Db 1535 SIPAKSPVGVDPVIVSHSPFDPHRRSSAAGEVYRSHL.PTHLDPAMPFHRALDPAAYLLO 1594
QY 1693 -----PS-YKFEVYKSTKKSIFVLIPNDLKKLARKGIREVPYFNNAKP--ALD 1739
Db 1595 ROLSPPTGYPQOYOLYAMENRTOTL--NDYITSQOMQVNLR--PDVTRGLSPREOPLG 1649
QY 1740 IMPYSPPTFGITWRYRLQTVKSLAGVSLMLRLMLASLRWDMAKVP-----GGG 1792
Db 1650 L-PYATRGIIIDLT-----NMPTIIVPHAGT 1676
QY 1793 STRETSSELTETTEIIRKRDVPGYGIREFYCIKKIICPIGVPEPKETPTPQKGLRSS 1852
Db 1677 STPPMDRITYIIGTQVT-----FPPRP-----YNA 1702
QY 1853 ALPRAPRPTPOTGVIIETWVAEELELWEIRAPAFERKEKAQAVBOAKKRLQOOP 1912
Db 1703 SLSPEHP-----THLA-----AASAER-EKEREREKERERERER- 1740
QY 1913 TVIATSTSPSTSTISPAQKVVAPISGSVTTGTAKMLVLTGVSPATVTPQONKNFH 1972
Db 1741 -----EBERIAAPADLYLRGSE-----QPRGSGHGV----- 1770
QY 1973 QTFATWVQOGSNGVVOYQVGLIIPS-----STGTSQOTFTSFOPRTATVTRPNTSG 2028
Db 1771 -----RSPSPSVRTOETIILQORPSVFOGNGTS--VITPLDP--TAQLRIMPLPSG 1817
QY 2029 SGGTNSGOVLTGPQIRGGMVITRPTLQOSTLGKAIIRTPVMVOPGAPQVMTQIIRGQ 2088
Db 1818 -----GPSISOGLPASR-----YNTAADALAL----- 1840
QY 2089 VSTAVSAPNTVSTPGQKSLTSATSTSIQSSASQPPRPOGQVYKLTMAQLTQLTQGHG 2148
Db 1841 VDAASAPQMDVSKTKESKHEAARLEENILRSRAVSEOOOLEQKNLEVE----- 1890
QY 2149 NOGLIVVIGOGQTTGOLLLPQGVTVLPFGQOLMOAAMPNGIVQRLFTPLATTATTA 2208

Db 1891 KRSVQCVCTSSALPSSKAQ--PHASVYSEAGKD--KGPPEKSRYEELRTGKTITTA 1946
QY 2209 STTTTVSTTAAGTGQORQ-----SKLSPOMQVHODKTLPPA-----OSSVGPAAQOP 2258
Db 1947 NFIIDVITROIADKDAFERGSSSDSSSLSHREYETASDALEVISPASSPAPPOEKFO 2006
QY 2259 TAOP-----SARPOPTQPSQAPQEVOTQ-----PEVOTQ 2289
Db 2007 AYQPDWAKNQAENESTROYEGPLHHYRSGQESPPQOQPLPSPSSQSGMGVPRTHRL 2066
QY 2290 TVVSHVPSAOPTHAOSKPOVAAOSQPOSNVQOSPVYRVOSPQTRIRPSPQLSPG 2349
Db 2067 ITLADHICQIITQDFARNQVPSQASTSTFOT-----SP---SALSSTPVKRTISSRYSP 2118
QY 2350 QOSQOVQTTQSPLP---IQPHTSLQIPSOQOPQOSQOPVQSSQTTLSSGQTLNOVSVSSPS 2406
Db 2119 SOSQ---TVLHPRPGPVSPENLVDSKSGSRPEKSP-----RSHIFS 2158
QY 2407 RPOLQIOQOPQVIAVAPOLQOOVQVLSQ 2434
Db 2159 EPEEIPSPGCP--AVHEKQDSMLLSQ 2184

Search completed: November 20, 2002, 16:26:32
Job time : 117.509 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:21 ; Search time 48.0411 Seconds

(without alignments)
5817.155 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAEETQDSE.....KIKGFKASRSHNNKIQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4082	27.3	810	2 G01252	small GTP binding
2	854.5	5.7	1711	2 T21432	hypothetical prote
3	546.5	3.7	5327	2 T13564	microtubule-associ
4	514.5	3.3	2447	2 T16870	hypothetical prote
5	490	3.3	5170	2 T15348	hypothetical prote
6	463.5	3.1	3507	2 T34513	hypothetical prote
7	462.5	3.1	3759	2 A35085	trithorax protein
8	458.5	3.1	3488	2 T34418	hypothetical prote
9	446	3.0	2897	2 B48666	cell proliferation
10	440	2.9	2271	2 F90073	hypothetical prote
11	439.5	2.9	3256	2 A48666	cell proliferation
12	439.5	2.9	3924	2 S37431	ankyrin 2, neurona
13	439	2.9	2938	2 T30249	cell proliferation
14	419.5	2.8	2232	2 T34434	hypothetical prote
15	417	2.8	2441	2 D71623	erythrocyte membra
16	407.5	2.7	3828	2 T13857	trithorax protein
17	406	2.7	4377	2 A55575	ankyrin 3, long sp
18	405.5	2.7	452	2 T21435	hypothetical prote
19	403	2.7	5105	2 T32650	hypothetical prote
20	402.5	2.7	510	2 T21430	hypothetical prote
21	401.5	2.7	405	2 T21433	hypothetical prote
22	401.5	2.7	1367	1 S48478	glucan 1,4-alpha-g
23	397	2.7	2722	2 T20532	hypothetical prote
24	397	2.7	6642	2 T29757	protein UNC-89 - C
25	393	2.6	1630	2 A53577	ascites stanolgyco
26	393	2.6	2481	2 D90011	fmb protein flmpo
27	392	2.6	1871	2 D96796	probable heat shock
28	392	2.6	2738	2 E88320	protein F07A11.6 l
29	391	2.6	1939	2 T18372	repeat organellar

30	391	2.6	3147	2 T18674	hypothetical prote
31	383	2.6	2346	2 T13829	rpr homolog - fruit
32	378.5	2.5	1229	2 T25697	hypothetical prote
33	377.5	2.5	2215	2 T16871	hypothetical prote
34	375.5	2.5	3259	1 A56539	giantin - human
35	375	2.5	2526	2 T20531	hypothetical prote
36	374.5	2.5	2116	2 A26655	myosin heavy chain
37	372.5	2.5	1589	2 T13606	mature-parasite-in
38	372	2.5	1526	2 A45605	giantin - human
39	368.5	2.5	3225	2 T52300	hypothetical prote
40	367.5	2.5	1829	2 T24583	nascent polypeptid
41	366.5	2.4	4187	2 T130826	hypothetical prote
42	366.5	2.4	2580	2 T18440	nuclear receptor c
43	365	2.4	2453	2 S60254	adenomatous polyo
44	362.5	2.4	2843	1 RBH0AP	hypothetical prote
45	361.5	2.4	2541	2 T29340	

ALIGNMENTS

RESULT 1					
G01252					
small GTP binding protein SEC4 homolog - human					
C:Species: Homo sapiens (man)					
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 29-Jan-1999					
C:Accession: G01252					
R:Browser, R. P.					
submitted to the EMBL Data Library, January 1994					
A:Reference number: G06441					
A:Accession: G01252					
A>Status: preliminary; translated from GB/EMBL/DBJ					
A:Molecule type: mRNA					
A:Residues: 1-810 <BOW>					
A:Cross-references: EMBL:U05237; NID:q1276427; PID:q451848					
C:Genetics:					
A:Gene: FAC1					
Query Match					
Best Local Similarity 98.7%; Pred. No. 3.8e-144;					
Matches 777; Conservative 4; Mismatches 2; Indels 4; Gaps 1;					
QY	1	MVSEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDDAASYCTESSFSRSHST	60		
DB	1	MVSEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDDAASYCTESSFSRSHST	60		
QY	61	YSTPGRKKRPRVHRPSPILKEKDIPLPEPKSSDLPMPNEHIMVIAIYELRNFTY	120		
DB	61	YSTPGRKKRPRVHRPSPILKEKDIPLPEPKSSDLPMPNEHIMVIAIYELRNFTY	120		
QY	121	LRSPRFEDFCALYSOCCTMAEMHYLLKAVIREEDTSMTTPGPAIDKQSVSTLY	180		
DB	121	LRSPRFEDFCALYSOCCTMAEMHYLLKAVIREEDTSMTTPGPAIDKQSVSTLY	180		
QY	181	FIDGMTPEVLRVYCESDKREYHNHLYQEAEADYPVGVENKIKVLOFLYDQFLITNIAE	240		
DB	181	FIDGMTPEVLRVYCESDKREYHNHLYQEAEADYPVGVENKIKVLOFLYDQFLITNIAE	240		
QY	241	ELMSEVIGYDQDCHRCVHKGLDLLCCETGSAYVHLCVCKPRPEEVPDEKQCVCAHKY	300		
DB	241	ELMSEVIGYDQDCHRCVHKGLDLLCCETGSAYVHLCVCKPRPEEVPDEKQCVCAHKY	300		
QY	301	PGVTDVAEIQKKPYIRHEPIGYDRSRKRYWFLNRLIIEPTDENKKIYYSTKYQ	360		
DB	301	PGVTDVAEIQKKPYIRHEPIGYDRSRKRYWFLNRLIIEPTDENKKIYYSTKYQ	360		
QY	361	LAELIDLDKDYAEALIKLIEEMREEIRHMDITEDLTNKAQSNKSFIAANEILIS	420		
DB	361	LAELIDLDKDYAEALIKLIEEMREEIRHMDITEDLTNKAQSNKSFIAANEILIS	420		
QY	421	IRAKKDDIDNVKSPEETEKDKNTENDSKDAEKREPEFQDSLEKSDSKTPDDPEQCK	480		
DB	421	IRAKKDDIDNVKSPEETEKDKNTENDSKDAEKREPEFQDSLEKSDSKTPDDPEQCK	480		


```

OY 1382 SEKKQORTSTPOINGKDKPKIYLKGECLKEISEKSVYSGNVEPKVNNINKIIPENDIKS 1441
Db 1002 -----
OY 1442 LTVKESAIRPFINGDVIMEDFERNSSSETKSHLSSDAEGNYRDLSTLPSTKESDSTQ 1501
Db 1002 -----
OY 1502 TTTPASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLNDPIDENGLPINKNEN 1561
Db 1002 -----
OY 1562 VNGEKKRTVITEVTMTSTVATESKYVAKVGKQIVVSTENCACKSTVTTTTVTYK 1621
Db 1004 VGEERK-----
OY 1622 LSTPTGSGVDIISVKEQSKVTVTTTVDLSLTGTGTLVTSMTVSKEYSTRDKVKLMKS 1681
Db 1010 -----
OY 1682 REKTRSGTALPSYKRFVYKSTKKSIFVLPNDLKLARKGIREV--PYFVYNAKPALD 1739
Db 1010 -----
OY 1740 IMPYSPRPTFGITWRYLQTVKSLAGVSLMLRLIMASLRMDMAKVPDGGSGSTRTETS 1799
Db 1063 IMPYAPRPTLDCWKMKQCTLNARSLHVAVALQKTIWSSIKFNEFD--DDTHPRDRVVI 1119
OY 1800 ET--ETTTEILKRDVGVYGRFECIKIKICPT-GVETPKETPTPRKG-----LR 1850
Db 1120 DTPSHDERRRILIRHEMPYGOYEREMEIEITPLYDEEEDDESLSNRKSGSSEFSHR 1179
OY 1851 SSALPRKRPET-----PKOT-----GPVILETVAEELELME----- 1883
Db 1188 SSSAKKKRRQRIEFLSLKGNVPKSKNAFRSLDNRAAIRREWDVGLKVEFLKDW 1239
OY 1884 --TRAFAEREKEKAQAV-----EQAKRLEQOKPTVIATSTTSPTSSTTSPAQ 1934
Db 1240 KMRLEAEKTAARKLEATRKAKAKEDERRRLOOQOQSVARIPV--PMS--LIPSE 1294
OY 1935 KMWVAPISSV--TTGCKMVLTKVGSFATVTPQN-----KNFQOTATWKKQO 1983
Db 1295 KMWV--PYLGSQOQRRPNRNGERFLEKYNSSSVSPQAHVASTPPPGYHQPQNIIRQAG 1353
OY 1984 SNSGVVQVQKVLGIIIPSSSTGSOOTFTSFOPRTAVTIRPMTSGSGGTTNSQVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPNFQS--RP-VATIIPTPOLRAAGADG--VIRAVM 1396
OY 2044 IRPG--MTVIRTPLOOSTLGKAILIRPVWVQPGAPQOVMTQIIRGO----- 2087
Db 1397 MTPGNKSTVNTSTSTPPQ-----ALNRQOYLOQOQOQPAVRRLTNGYHFMQDMRGGR 1451
OY 2088 -----PSTVASANVTYVSTPGOKSLTSAT-----STSIQSS 2120
Db 1452 NPSVQMHQRLPQNRALQRFGESESTTEMRVTEALIPNDGDGQPPVIRPYDPTSFDAQ 1511
OY 2121 ASQPRPOQOGVYKLMACLTOLTOGHGNOGLTVVIOGOGTTGOLQ-----IPQSVTVL 2176
Db 1512 RAOQOHPOGRPVYSTPAQIKRTTQPGVKH--NVILAKMSDGTQKVVLPKPGFPEPTVI- 1568
OY 2177 PGRGOQLQMAAMPNGSTVORELFTPLATTAATTTTSTTAAGTGEQROSKL----- 2230
Db 1569 -STGGRVIVYRQPTAVOQOLVY-----ATPGTRVVRIPNANGAPRQOQDHQVMRVVO 1621
OY 2231 --SPQKQVHQDK--TLPPAQ-----SSSYGPKAKQPTQKQPARQOPQIQPOS 2274
Db 1622 ASQPRAMEYMDQGTTPPQOQVRYVLOGGNSGTPTVNPVPEKVSRRGPRGLTMQVVOQOQ 1681
OY 2275 PAQPEVOTQPEVOT--OTTVSSHVPEAO 2301
Db 1682 QHNPAHYMDPATGFAVSTTTTQVYDDEQ 1711

```

```

RESULT 3
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
A:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match          3.7%; Score 546.5; DB 2; Length 5327;
Best Local Similarity 18.3%; Pred. No. 4e-12;
Matches 550; Conservative 479; Mismatches 1154; Indels 825; Gaps 121;

OY 2 VSEEEEDDGADEETQDS-----EDDER-----DEMEEDDDSDPEEMEDD 44
Db 2191 VKDEAKQSKRESEKTESVTPKAKDKSPKVLQPVSTETIRREDADQPMKPSQASRR 2250
OY 45 DD-ASYCTESSFRSH-----STYSSTPGR-----RKPRV-----HNRSPFILE 82
Db 2251 ESIAESIKASPRDEKSPASKASRPGSAESIKYDLKPKQIKDKDSTHSRRSLED 2310
OY 83 KDIPLEFKKS--SEDLAVNEHNMVIAIYEVLRNGVLRISPFREFD-----PCALV 136
Db 2311 KSAVTSE--KSVSRPLVASDH--EAAVALEDKAS-----STSP--KKSRRGFALFTV 2359
OY 137 SQ--EECTLMAEHWVLLKAVLR-----EEDTSNTTFGPADLKDSVNSTLYFTD 183
Db 2360 SSRIEATMEFKIEVEKSSIALSLQGGSGGKLTQDSSPVVAEEDPFAVAASVSTVP 2419
OY 184 GMTWP--EVLRVYCESDKEYHNVLPQOEAEDPYGP-----VENKIKVLIQ 226
Db 2420 TLTKPAELAQIAAKTVS-----SPIDEALRTSPAPHISIRADSPACASEIETASQKSPQ 2475
OY 227 FLYDQFLTNIRAEELMSGVLIQYDDHCVRVCKHGLDLCFCFCSAAYHLECKVP----- 280
Db 2476 VLKSSRPAMVAESK-----DDAQLKSSVEDLR-----SPVASTETISNPASAGET 2521
OY 281 ---PLEEVEDEMOCEVYAHKVPVGTDCVAETQKKKPYIRHEPIGYDRSRKKYFLNRR 337
Db 2522 ASSPIEAKRDAEFP--QAEK--AVLPRLTELKGLP----- 2555
OY 338 LIIEEDTENENKIKIYVSTKVQLAELIDCLDKYWEALCKILEMREIHRHMIDTED 397
Db 2556 -----TLSPVDVAHASVQPAELSKV-----DIEKT 2581
OY 398 LNKANGSKSPLAANEELIESIRAKKGDIDNVKSPETEEDKNTENDSDAKENREE 457
Db 2582 ASSPIDAEKSLIGPAE-----RPESPASADDAASVKSADKASPPSV 2628
OY 458 FEDQSLKESDD-----KTPDDPPOGK--SEPTFVGDKGNSVANT--GDNTTNAV 506
Db 2629 VESTKADSTKGDISPESVLEGPKDDVKSKESSRP-----SVASITGDSTKDV 2681
OY 507 -----SEETSPSEGRSPVGLSETPDSSNNAKKVASELPQVDEEENKTCESN 556
Db 2682 RPASVYESVKDEHKAESRESIAKYESVIDEAGKDSKSSS---QDSQDKESTLASKE 2738
OY 557 TSATTTTSTQPNLENSNSSELSNSQSESAAKADDPENGRESHTPVSIOE-----EIVVD 611
Db 2739 ASRRESV-----ESSKDAEKSESRESVYASGEPV--PRESSPLDSKDTSRPGSVYES 2792

```


[illegible]

Dd	366	ELSDNEDVNSKRCSCSTSKKTTNRSDSSDSDNDSDRDELQKKRRKM-----	KRKNVPTKD	419	
Qy	499	GDNTTNATSEETSPSEGRSPVGLSETPDSSNNAEKKVASELPDQVPEEPPNKTCESSNTS	558		
Dd	420	GSDVSNESFEDDASGE-----	VVATKL--IKESKKRRCPRPKK	456	
Qy	559	ATTTSIGONLENSMSSELNSOSSOESAKAD-----	DPNGCRSEHTVTSIOE	606	
Dd	457	FAPELITBEDITPTPEDSITJSSDERDNDADPYAIFYQKEFNRP--	RRDH-----	506	
Qy	607	EIVGDFTSSEKSTGELSESFGAGKASGSTRITRLNPNPSKLSQOLKSQOQVAAAHEANKL	666		
Dd	507	-----PEKD-----	KLVNPFMDIDM-----	NHQVPRK	528
Qy	667	FKEGEVLYVNSOGEISRLSTKKEVIMKGINNYFKLQGBGTRVYHNQYSTNSPALNKH	726		
Dd	539	FEKGEITHIVISDDNSSEHSEDEAE-----	DRSSSDIS-----	560	
Qy	727	QHREDDHCR--RHIAHFCFLTPAGEFKWNSVHGSKVLSTLRLITLOLENNIPSPFH	784		
Dd	561	-----EHEKEISKEFLSHQPLRPNITSUDDQV-----	ITVAKD-----	595	
Qy	785	PWMAASHRANWIKAVOMCSKPREPALALALECAVKPVMPLPWREFLGHTRLHRMTSTER	844		
Dd	596	-----VKOASATSKP-----	GETSPDSSKIE-	617	
Qy	845	EKEEKVKKKEKKOEEETMOATVWVKTTPVKHQWKKQKGEERYTGVCSGMWIMKITHY	904		
Dd	618	EKPDV--SEEVSDMEETPEHTIADKGTDFLNNIHEDEM-----	YGF-----	663	
Qy	905	REVPALPGNTVNVNVRKSILEGKNNMENDMESDKRCSRPKKIKIEPD-----	SEKDEVK	960	
Dd	664	R-----PDGTGIS--RPVYQAPGPNRLSNMVCPEAVLKGKKKVLVLPEDYEISSDERYT	717		
Qy	961	GSDAKGADQNDENMISKITEKKDDVKELSDSDKPCKEEPMEDVDDMKTSSH-----	10144		
Dd	718	LSDSEETSPSAEMQOSETSEAGPSTIIKTSGTERTQOGSSPPEPSTSDRHMHRKIDT	777		
Qy	1015	-----VNCQESSQDVNVNVEGFLRT-----	SYKKTKSRLDGLERRIKO	1057	
Dd	778	RRRKIADSDSDLPDVEYSIDGNELVATGKPIIKHKFYVSANRMPKSNULDFTTGRNARE	837		
Qy	1058	FTLEBKORL-----	EKIKLEGIKGIGTSTNNSKNLSESPVITKAKEGCOSDSMOE	1110	
Dd	838	IPMEISIKRLAEQVAHEEYKIHRRQVYLEAVEAASKLNVYVDTTEEEE-----	IEE	891	
Qy	1111	OSPANNNDQRPDLQOGSOSQSDSVLHRSDSHTT-----	NKLVPRQVLDVDSIRSPERK	1166	
Dd	892	ETPE-----EEVVK--VASPTAIATENTJTSTABPEEGVAMKEPIEITF--	DPDEPC	942	
Qy	1167	PKONSIENDI-----	EKVSDLASRQGEPEKSKTKGNDFFIIDSKLASADIGITLIC	1218	
Dd	943	SSAQAOQHELIIHEVKEQOIIEDSLQNNKPS--SKT-----	VKSESEREA-----	987	
Qy	1219	KNKKPILQOESDITVSSKSKALHSSVPKSTINDADATPLSRADFEKGLACDSESNSTLEN	1278		
Dd	988	-----QEERLEKDMESEQOKKADNFWTEVDKESSEASSSESDKSDFE-----	DETIDA	1036	
Qy	1279	SSDPTVSIODSSSEDMITVQNSNESISQOFTRR-----	EODVEVLBPILKELCVSGESTGNC	1332	
Dd	1037	QSKTVKISLKHKEKV-----	SDEEL--EDFTQKPEEPAVATADAKMKIRITIGEYSTEFL--	1068	
Qy	1333	EDRLPVKGTENGAKKPSQOKKLEBRP-----	VNKCSDQIKLKNPTDKKNENRESEK--	1384	
Dd	1089	-----KLVAQOQPAVYDDEVYALGFCVRNMDQEFSTIKETGKTKNKPDESBV	11359		
Qy	1385	KGORSTFQJINGK-----	DNKPKITLKEGCLKETSERVANSVGNVEPVNINKIIPEN	1437	
Dd	1136	KNFEES--FAAKHRVPRPKLPNITERKTI--	ERAHNMYKYHVV--DMEPLHKILIALQKO	1190	
Qy	1438	DKSLTYAES--AIRPFGADVIMEDFERNSESTKSHLSSSDAEGNVRDSELTLPST	1494		
Dd	1191	QIATCANLISOPVYVPEEAEQOVOLLHNQNSILRPL-----	NNPOFALILHAAO	1243	

```

OY 1495 KESDSTOTTPASCPSPNSVNOVEDMEIETSEKKVYTSPI-----TSEESNLNDP 1548
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1244 QQAIOQORAOQKQOTOKELARQAEQARVELAKRRIQAQDAEKALROKEQMSNVS--- 1300
OY 1549 IDENGELPINKENVENGESKRVITYE--VTMTSTVATESKTYI-KVEGDKQ-----T 1599
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1301 ----GIVSSDQNAQSSNAQOTGLIEMQTTTJNSDSLIRPNTLADNSHLCESQOIPVIES 1356
OY 1600 VVSTENCASST-----VTTTTVTKLSTPSTG----- 1628
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 IQSTSEALKESBENYKMDPILTPASTVSSKSAAPATRRPRPCSSYDRPSSPSVIRERL 1416
OY 1629 GS-----VDISVKE--QSKTVVTVTDSLT-----G 1656
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1417 GSDGALINRPNCNDKSRSPISRAPEYETIRINDHGNETHILAGNITHYETILL 1476
OY 1657 GTLV---TSMTVSKEYSTRDKVKLMKFSRPKTRSGTALPSYRKFPYKSTKKSIFVLPND 1713
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1477 GTSIGQDSTRYDQEGSTQYID-KTIDDNSKNGTNDVEQSNVL-----LREN 1526
OY 1714 DLKRLAKGSGIREVPYNYNAKRALDIMPSPRPFGITWKRRLQTVKSLAGVSL---- 1769
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1527 DLNR-----EMLRAN-----RHPSTMLAMGNLSINERH 1556
OY 1770 -MRLMLASLRMDMAKVPDGGGSTRTETSEITTEITTEIKRRDVGPGYGFREYCIRKI 1828
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1557 NKVOQVLAISELODLTAKHSGA-----VSQTVQV-----VOGEG----- 1592
OY 1829 IC-----PIG-----VPETPKETP-----TPORKGLS----- 1851
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1593 VCAGTSDAIGETDEDDVEEPEFTVDQLAKKILKROGLESESDSDDEDMDVYDND 1652
OY 1852 -SALR--PRRPETPKQGT-----PVLEETVAEELELMEIRAAEVEKEKAQAVE 1900
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1653 GSVIRRAPHKKRETRKKKNTFVPNIPKIRKTV-----DKKIEKERA--- 1695
OY 1901 QQAARLEQOKPTVIATSTSTSTSTSTSPAKVM-----VAPISGYTGTGKWLTT 1955
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1696 -KRRARKSKQKMASIRIAY--PQKPTQGFATPQQVPGKPKHSAAMAAARATPRKPKKMS 1752
OY 1956 KVGSPATVTPQO--KNFHQTEATWAKGQSGNSGVNOVOQKVLGIPTST--GTSQGTFT 2011
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1753 NVSPATKITATVPVNAAGFHQ-----NQOQLSDMAQ-----POSTPIRTTPOGT 1798
OY 2012 SPQPRATVTRIPNTSGSGTSSQVITGROIAPGKTVIRTLQOSTLKAIIIRPVNV 2071
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1799 GSAFOATPOSHLAQLQGFVNGAMQ--QAPQ-QQGM--YTAAQLQAMQAAVAQTAQAA 1852
OY 2072 Q-----FGAPQVMTQILIRGOPYSTAVSAPNTVSTPGOKSLTSATSTSIQSSASQPPR 2126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1853 QAAYAAEAATQAQVQAARAPRQQLVQROVYGH--PGQVNVMPRQMMLN-QGNPQAVN 1910
OY 2127 PQGQVYKLMAQLTQLTGHGNGOGLTVIQQGGQTTGQLDLPQGVTVLPRGQQLMDA 2186
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1911 PAQAQMDDERKKMEV-----NAVYHMQSGH-----PPRNQELFYQ 1949
OY 2187 AMNGTVQRLFPPLATATATTTTSTVSTAAGTGEOROSKLSRPMQVNOHQKTLPPAQ 2246
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1950 QF--AAQADLRSAAADAQAQAQAQAOATMRRQAQAEAVAAQAAMKQEOAR-----AQ 2000
OY 2247 SSSVGAFAKAPOTAPSARBPOTQPOSPAQPEVOTQREVOTQTVSSHPSAEAPTHAQ 2306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2001 AAKKEAARLKAETEAkakakQAQAEARAKAEQEMRVR-----QAQAQOTQ 2045
OY 2307 SSKPQVAAQSPQSNVOGQSPVAVQSPQTRIRPSTPSQLSPQOQSOVOTTTQOPIDIP 2366
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2046 AAQAQASQAHAAQNAQOTQATLVEIORMIQ-----SCQPLSMQ 2083
OY 2367 HTSLQIPSQQPSQPVQVQSTQTLSSGQTLNVY-----VSSPSRQQLQIQPQ--- 2417
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2084 MQLQMQSVQVQOMQAOQVQOQMQOQLOMQQFAARMQOGTEPRPAVSQAQVQOGMPAGT 2143

```

```

OY 2418 QVIAVPOLOQOVVLSQIQSVVAAQI-QAQQSGVPPQIKLQPLQIQSSAVQTHQIONV 2476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2144 QGMPTGGMQQLQGLGIPGNGQLPQAGSQSQTSQAQOQQQLQFMQLQIQQQQLMQ-HQLQQQ 2202
OY 2477 VTVQAASVQELQORVQQLPQOQKQKQOQIEIVNVPKSKLLIVEIT-QQVYMKHNAVY 2535
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2203 LQMQHQHQHQOQOQOQOQIQMQ-QOQOQLAQOGLVYPVSS-AWLQOQALAQOQOQVQOQNL- 2259
OY 2536 EHLKQKSKMP 2546
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2260 --MKVPSAQTP 2268

RESULT 5
T15348
hypoethetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15348
R:Gatung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: EMBL:050071; NID:g1208871; PID:g1208877; PIDN:AA93447.1; CESP:B0
A:Gene: CESP: B0350.1
A:introns: 48/1; 5039/3; 5116/3

Query Match
3.3%: Score 490; DB 2; Length 5170;
Best Local Similarity 17.3%; Pred No. 4,7e-10;
Matches 640; Conservative 540; Mismatches 125; Indels 1274; Gaps 162;

OY 1 MVESEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 LVSEHEHHPACDSERL--SEPAQSPPEVETHAESQSNLVEITTTTA--VTRFYDEQGE 621
OY 61 YSSTPGRRKRRVRRSPRIIEKNDIPLEPKS-----SELMVPNEHIMVIAIYELVN 116
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 EQASSRATKERIE--QSPVASERSIVSTENRQSPQSESESLPTSE-----KN 668
OY 117 FGVTLRSLPRPFDFCAVYQOCCTLMAEMHVLKAVIREEDTSNTFT---GPADLK 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 VHTVTEITTYTRCFEPIASF-----LEHASREFEQGSNDPRRSVSPVDQA 715
OY 173 DSVN-----STLYFDGKMTPEVLRVYCESDKYHNHVLPRYQAEQY 213
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 716 EQVDVHPSPAPSSHAESQGVPSPHIVETTTTTVTRTF--QDEEYPR--PESPAEIT 771
OY 214 PYGPVENKIVLQPLVDQFLITINARELMSQVIOYDDHCRCRCHLGLDLCCEFSAYV 273
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 PIPSSQOSEPHIVKETTITTTTVTRF-----LYDE-----PKGNTFSPADSS-- 816
OY 274 HLCGVPRPLEEVEDEDMQCEVCVAHKVPVTDCAVLEIQKRPYIRREPIYDGRSRKRYMF 333
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 817 HAESR---QVPE-----SPVYSIQEYH----- 837
OY 334 LNRLLIEEDTENENKRIYVSTKQVLAELIDCLDKDYAEALCKILEEMREIHR-HM 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 838 ----VETTTTNVTSNITDDEDNVSSB--DPATQHFQOSE-----TSVHRSHP 881
OY 393 DITEEDLNKARGS-----NKSFLA--ANEELIESIRAKG-----DIDNVMS 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 882 DSVSESDGEGIGSKVIGFAKKAQAGVAAVAAVAAVAAVAAVAAVAAVAAVAAVAAVAAV 941
OY 434 PEE--TEKQKNETENDSKDA-----EKNEEPEQD----- 461
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 942 PESPVPETQYQDDSAQSSHTDFEHMHPESP1EHKETEEDHSHSPESVLSKEKEH 1001
OY 462 -----SLEKSDSDKTPDDDEQK-----SE 482

```


QY	2118	QSSASPRPPO--OGQVYLTMAQLLTQLQCHGGAGGLVYVYQGGTQTQQLDLPQVYV	217
Db	3074	DPAAASVSEEDVHQIQ--TTTTTIVYTRH-----VPEEDIDSGMELEKYSSE	3122
QY	2176	LPGPQQLMQAAMPNGVYQRFLLPLATTTATSTTTVYSTTAAGTGEORQSKLSPPQO	2235
Db	3125	SPVPSSE-----DSNVLETTTTTVS-----	3144
QY	2236	VHODKTLPPAOSSSVGPAAKAPQTAQPSANRPQTOQPSAPQEVOTQTOPEVOTQTVS--	229
Db	3147	--REHFPEEDHSHV--VESQETYSAGSP-----VPSEKSVPRVLETTTTTVTRE	3193
QY	2294	-----SHVPEADQPTAAQSKCPQVANAQSPQMSQOGSPVRQSPQSRIRPSPQSL	234
Db	3194	HFDTEEDYTPESKTRSHDDGTTDD-----HVSQSPV-----PSE-E	3222
QY	2348	PGQSQVQVOTTSOPRPIQPIQHTSLQIPSQGQPOSPQVQSTQTLSSGQTLNQSVPSPR	2407
Db	3230	DDHVTHEQIDINDPRLDQVYSHSKYKE--SSIPSQDSTHYI--ELVMTSPITS---	328
QY	2408	PQLQIQPQPOVYAVPQLQOQVQVYLSQISQVYQAQIQQAQSGVPQIKQLPQIQQSSA	2467
Db	3281	-----ERYDPEV-----EKDVESADEI-----DSSAQYKXKSESVQTEKSL	3320
QY	2468	VQTHQIQN-----VTVQAASVQQLORQOQL	249
Db	3321	LAKQOQEGSDSDSEGSKYLGFAKAGVAGCVAAAPALAAVQAKAAYALAKDKDDE	338
QY	2495	RDQQOKR-----KQOQIEINVTSPKLLIYV-----IIOK	2522
Db	3381	EDQEDKEPFLIGFHKDDPDISQEPSSAVDSETPQATFYEPREEDKVITDASDSVQD	344
QY	2526	Q--VVMKNNAVIEHLKQKSTPAERENQRMIVCQWAKYIILDKIDKEKPAAKR--	258
Db	3441	EPKTIVPYDSTPEHNSN-----DREEPSIV--KSEGPYIVESTDAQTSAPSPRIS	3491
QY	2581	-----KREESV-----EQRS-----	2591
Db	3492	PVHSDAGASSFKRPESVTGEDEKNALPEISFSDAPYVIDSEYFENNNDQGRISPAHS	355
QY	2592	---KQNAKTIALLEFKHKSQLEIITKKRALLDKQLOI-----	2626
Db	3552	DEEDENDAEVIDSFHYHSQEOUNNE--EDPSIYSESEYISGHSRPREDSSTTVLVN	3609
QY	2627	-----EVOELKRDJIKKENDQMLQAGATVAAPCRP--VTPVLPAPAPRPS	2674
Db	3610	HHEPALPEPVPDE-----ELQDSRILIE--SEYKTSPLPPTSYTVVHEHPAE---	3659
QY	2675	PPPGVOHTGLSLPRL--EVASQOKRREEK--DSSSKSKKKMISTSKTKKDTKLY	2730
Db	3660	-----IHKYNTSTPTIVTVVSEHLDNRQNEPVPVSESEYTNASPLGPRPSPGSGPL-	3712
QY	2731	CICKTPIDESKFTYICDRCONMYTHGRCVGLQSEAEFLDEYVCPQOQSTEDAM-----	2784
Db	3713	-----PREEDSH-----VIESH-----EYTSPPVS-EDSVKHAVIEKT	3745
QY	2785	VLTPLTEEDYGLKRLVRLSLQAHKMMAPLEPVPDAPDADYGVYK-----E	2831
Db	3746	TTTTVTEERYE-----PEDSH-----SPPSBD--DVNGFVKYTTTTTTTTVTHNEPE	3789
QY	2832	PMDLATMEERYQRRY-----EKLTPEVADMTKIFPNCR--YVNPSPD	2871
Db	3790	PEHSTDBEHVESERYASGSPVPSSENSNNVETTTTTTVTTRHNEHPED	3838
RESULT 6			
T34513			
hypochemical protein zk783.1 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999			
C:Accession: T34513			
R:Favella, A.; Vaudin, M.			

submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GND0021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Intons: 14/1, 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 3.1%; Score 463.5; DB 2; Length 3507;
Best Local Similarity 19.0%; Pred. No. 2.8e-09;
Matches 444; Conservative 282; Mismatches 857; Indels 755; Gaps 92;

QY 419 ESIRAKKGIDNVKSPETEKKDNETENDSKAEKNRE---EFDOSLEKSDDKTPDD 475
DB 997 ESSSISTGKSGSKG-----PADKSDKNVPKGDXNDPDTTGDGEISTETSGEGGP--- 1048
QY 476 PQGKSE-EPTVEGDKGNSVANIADMTNATSEETSEGRSPVGLSETPDDSSMAEK 534
DB 1049 --KGKSKGQP--GDGKSEVK-----KPTSEVDGP-----GNISGTRKSNVPLK 1089
QY 535 KYASELPOVPEEPKPKTCESSMTSATTT-SIOPLENSSSELNNSOSE-----SAKA 587
DB 1090 -----PTDLPESGILITSSGKNSFTFEGTKLERLPKTEKDSSETPOLGLEISAGK 1143
QY 588 ADPPENGRESHTPYSIOELIVDPTSEKSTGELSESPGAGKAGSGSTRITRLRNPD 647
DB 1144 KPEPEEG-----TSKEVGLEILWESTPGSTLTDSDVGLETISGDLTK-ATKKPVEIE 1197
QY 648 LSQLSQVAAAHANKLFK-----EGKEVLVY-----NSGELISRL 685
DB 1198 GSGTDELTATTRVSKSTKRPVYVDGNGETSGVDGKPTTPAPRPSSSASTSRI 1257
QY 686 STKKEVIMKGINNFKLGQEGKYRVVHNOYSTNSFALNKHQHRHDHDKRHLAHKFLT 745
DB 1258 PTTSEASPEBG-----SGEAGVPESPDSGESHSTA----- 1288
QY 746 PAGEFRKMGSVHGSVLTISTRLTITOLENNIPSPFHPWASHRANWIKAVOMCSKPR 805
DB 1289 PDG-VSPPTSSATAPAPVPTTSASTPDVAEEGIPST-----SKPT 1337
QY 806 EFLALAILIECAVYKRVVMLPIRREFLGHTRLHMRSTEREKEKYKKKEKQEEETMQ 865
DB 1338 AEPLETTAPSTEV-----TSPGSGTEESTLPTPTGSGESTTSS 1366
QY 866 ATWVKTTPPVKHQVKKQGEYRVLYGQWSWISKTHYRVRPKLPNGTNNVYKRSLEGT 925
DB 1367 APTVEPAVTLPPQNNEKKEPEPTKDT-----FALPTTTTGAPQAN-DSVVENT 1412
QY 926 K-NNMDE-NMDESOKRKRSPKTKIIPDSEKEVKS-----DAAGADON----- 971
DB 1413 KCTSSDEGLD---ALCERRTGYCCEPGEAGAPPKSCVVDVDECATGDNHCHESARQ 1468
QY 972 -----EMDISKITEKKDQDVKELDSDS----- 994
DB 1469 NVYGVYACFCPTGFRKADGSCQDIDECTEHNSTCCGNAMCAVMPGYSCECENGLGD 1528
QY 995 -----DKPCKEPEMVEDDMKTESHVNCQESS---QVDVNVVS-EGFHLRTSYKKKT 1042
DB 1529 GYQCVPTTKKPC-----DSTOSSKSH--CSESNMSCVDVTDVSGVECKCMGMYKSG 1579
QY 1043 KSSK-----LDGLLERIRIQ-----FTLEBKQ-----RL 1066
DB 1580 KYCEDINECAVEKAPCSINAMCVNNNGTFFSCSCQGYRGDGFMCTDINECEDERHPCHPHA 1639
QY 1067 EKIKLEG-----GIGIG-KTSTNSGKMLSE-----SPVITAKREG----- 1101

```

Db 1640 ECTNLEGSFCKCEHSGPEGDGIKCTNPLERSCDEYKFCGRVDHVSCLSVRIYNGSLSS 1699
QY 1102 -CQSD-SMAQESPA-----NDQPEDLT----- 1124
Db 1700 VCECEGFRFKSNSCVDIDECESRNNDPASAVCVETGSGYRCEAGYEGGVCT 1759
QY 1125 -----QGQSQSSVLRM-----SDPSHTTN----- 1145
Db 1760 DIDECRGMAAGCDSAMCMCKRMGSCGCKMAGTGDGATCIKIEEPKSDKTADEMSR 1819
QY 1146 -KLYPKRDVLVDVSIKRPETKCPKONSIENDIEE-KVSDLASRGE-----PT 1191
Db 1820 LCELEKKQCTVDEEVRPGGACLPGHHPINGCTQSLQISLCKQKNDCKNHACIDIHED 1879
QY 1192 KSKTKNDPFI-----DCKLASADIGITLIC-----KMKPLI 1225
Db 1880 SHKSCPDGFIQDMGICDDVDCECNMAGMCDDENTKCENTGFNCVCLGEGFKKYDEKCVY 1939
QY 1226 QESDITVSSSKALHSVPKSTNDADATPLSRAMPEGKLGCDSNSTLEMSDQVSI 1285
Db 1940 DEKKOP--NREKIEIDENSSSSGQEKPTKGI--VSSSTATSESTTAEPHVTTSSIS 1995
QY 1286 ODSSEDMIVONSNEISISEQFTRQDVEYLELCKELVSGESTGNCEDRLPVKTEANG 1345
Db 1996 STSTDMTSSKSPENVMTSSSEPEVST-----SKSTTASETTIVSSTPSESS 2044
QY 1346 KPSQOKKLEERP--VNKSDQIKLNTDKKNENRESEKKGQRTSTQINGKDNPKI 1403
Db 2045 --SSAPLITSPATTEVTTESSVSKTTPKESSESEIVKLSSK----- 2087
QY 1404 YLKGELKEISERVYSGVNEPKV--NNINKITPENDIKSLYKESAIRPFINGDYMED 1461
Db 2088 -----PEVTESSVSSPSTPTSTQSVSTVPEKSKSVLSSEAPVSTPTEV----- 2136
QY 1462 FNERNSSETKSHLLSSDAGNRRDLETLPTKESQDSTQT----- 1503
Db 2137 --HTSEERKPLSASST--GDNSTTTPSSLSASKSTAPGCTASAVPVLSL 2189
QY 1504 TPASCPESNSVNOVDEMEIETSEVKVYSSPTTSEESNLNDFTDENGLPINKENYVA 1563
Db 2190 SPVSOQPTKTFEPATSSVQASSETSSGTSVKSTSEPSHVTKLSTSS--NPSSVAP 2245
QY 1564 GSKKRTVITEVTMTSYATEKTIYKVEKDKQTVVSTENCAKST--VTTTTTV 1619
Db 2246 VTSKPTPTVPESTEQPTTSPGOSLTPMNSNEVLTTSEPVLTSSLSLSPVSOSSSTP 2305
QY 1620 TKLS-----TPSTGGSVDIISVKEQSTIVTTVTDLSLTGGLVTGWSVKEKSTND 1673
Db 2306 NNLSESTIVETPKTSEVSLNBEPESTTAPPTLSPDILSTTNNLSQSTV--STED 2361
QY 1674 KYKLMKFSRPKRTSGTALPSYRKFYTKSTKSIFFVLPNDLKLARKGIRE-VPEFVY 1732
Db 2362 RSBISEENSEKPTSAPELVTSVTHVASSPD--VPRESSPDULLGSSSTNIP--EA 2415
QY 1733 NAKPALDIPYR-----SRPPTFGITWRYRLQTVKSLAGVSLMLRLMLASLRWDMAAYV 1787
Db 2416 SSKQTISSPTPTDPTTASSEPKTSTMSPDLSSTSNVLSSES-----ST 2458
QY 1788 PPGGGSTRTESETE---ITTEILIKRDVPGYGRFECYCIKTIPTGVPETPKETPT 1843
Db 2459 TPESSSKSVSSSTBGISVVTSTERSK-----VPS----- 2489
QY 1844 PORKGRSALRPKRPETPKQGTGVPVLIETVAEELE-----LMEIRAFAEYV 1891
Db 2490 -----TISVLEEDLTKT--TPSPILEETTTASETSEPLEDLSLTVSVAILHETTSSENV 2542
QY 1892 EKEKQAVQOAKKRLQOKPVIATSTTSPTSTSTISPAOKVAVADISGSVTTGTVM 1951
Db 2543 PKE-SESTTSSSESKPQSEPAGITLSTIVVPTSSVSLITASB-----IEATISNTPF 2594
QY 1952 VLTTKVSPATVTFQONKHFQTFATWVKOGOSNGVVOQKVLGIIPSSGTSGOQT 2011
Db 2595 -----KQKRT-----PITTSKSLVKST--TSPSTIV 2619

```

```

QY 2012 SFQPRATVTVIRPNTSGSGCTTSNSOYITGPQIRPGMTVIRTPLOOSTLKAIRPVWY 2071
Db 2620 SSEPSESTYK-----TIVSTIVST-----TPTTEETTSSESLIT--A 2655
QY 2072 QPAGPQOVMTQIRG--QPVSTAVSAPNMYSSNPQOKSLTSAFSTNINIOSASQPRPQ 2129
Db 2656 APSKPTSESTSESEAPPTPKTSETKPSNVSS-----TSRKTTEVEESTSQ----- 2702
QY 2130 GQVKLMAQLTQLTQGHGNGGLTVVIOGOGOTTGOLILPQGVTVLPGGQQLMOAAMP 2189
Db 2703 -----SGSLF-----STMS 2712
QY 2190 NGTVQRFLETPATATATATSTTTTV-STTAAGTGQRQSKLS--PQOVHODKTLPPA 2245
Db 2713 STSEPEETNAPAVTVASSEASTTLEENSTSPSSSEASVKLSLPESITSEAVTV--- 2769
QY 2246 QSSSVGPAKAPQPTAPPSARFPQPTQPOSAPQEVOTQPEVOTVSSVSEAPPTHA 2305
Db 2770 --SSRAPAEI--TMSSESHREISTVSEPESEIPLSTTVSPNVYASSIPSE----- 2818
QY 2306 QSSKPVAAQSOQ--QSNVQOSQEVVQSPQRTIRPSTPSOL-----SPQOSQOVOTTS 2359
Db 2819 -----EPILSVYSSSTPRV-----RLTGTPEDDLIVSVTVPSHGKRRQNTA 2861
QY 2360 QPPIQPHTSLOIPSOQPOQPOVOSSTQTLSSGQTLNOVSVSPSRQLOIQOP 2417
Db 2862 SSVPSNSTSPILPSSLSLTPQGP-PTTTTAKPATTSGKRGSPSIQPAEMFTTPAP 2918

RESULT 7
A33085
trithorax protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
C:Accession: A33085; A38240
R:Maizo, A.M.; Huang, D.H.; Wozer, B.A.; David, I.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
A:Title: The trithorax gene, a trans-acting regulator of the bithorax complex in Dros
A:Reference number: A33085; MUID:90192757; PMID:2107543
A:Accession: A33085
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3759 <MAZ>
R:Maizo, A.M.
submitted to GenBank, January 1990
A:Reference number: A38240
A:Accession: A38240
A:Molecule type: mRNA
A:Residues: 1-2361, 'Y', 2363-2397, 'N', 2399-2405, 'N', 2407-2411, 'N', 2413-3759 <MAZ>
A:Cross-References: GB:M31617; NID:9158817; PID:9158818
C:Genetics:
A:Gene: flyBase:trx
A:Cross-References: flyBase:FBgn0003862
C:Key words: Drosophila trithorax protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 3.1%; Score 462.5; DB 2; Length 3759;
Best Local Similarity 18.7%; Pred. No. 3.3e-09;
Matches 627; Conservative 422; Mismatches 1151; Indels 1157; Gaps 156;

QY 5 EEEEDGDAAEETQDEBD---EEDEMEEDDDSDVPEMEEDDDDDASCTSSFSRSHST 60
Db 538 DEDEDEGVFFRRNDSPEDQNNMEDDDDDDEFEAEEDQNDQNDDEAASKSAETERS 597
QY 61 YSS-----PGR--RKPRVHRPSPILEKDIPLPEPKSSEDLMPVN 101
Db 598 AGADERDPDEKOLVMDSHFVLPKSTRSSRIITKPNKRLLEBAI-----STKKPLSLGD 651
QY 102 EHMIVIAI-----YELRNFGN-----VRLSPFRE 129
Db 652 SKGKNVFGTSSSAGSTASTFASATNLKLGKTFEFNFGTLKRNSSAAGIYFLRQPLQ 711

```

QY	130	DFGALVYSQECJTLMAEMHVLKAVLAREDDNSNTFFGADLKDSVNSTLYTIDOMTPRE	189
Db	712	-----ADNOQATFAAPKACPTSPSPATIPKPAANLASSFGSLASTNSTRV-----TPPSS	760
QY	190	VLRAYVCE---SDKRYVHHVLFYQEAEPYRPGVPENKI--VLFLOFLVDOEFLTNAREELMS	244
Db	761	ACSI--CAVVVSSKRV-----TQARKYGVAVADVCRRFFSKMTKKTSANSTRAN--TS	810
QY	245	EGVYQYDHCRCVCHKLDDLIC-----CETGSAYVHLCVYKPLEEYPEDEMOCEVC--	295
Db	811	SGSOOY-----LOCKNEGSPCSIHSAKSQLNFKRFFKYD-----RCTACWL	852
QY	236	-----VAHK-----VPGVY--DCAVAELQXKKPIRINREPIGYDRSKRYVFLNR	337
Db	853	KKCMISFQOLPAHRSRSLATILPPGMMGEAAAREKSAEILL--SPGSLCF-----	900
QY	338	LIFEDTENENEREKIWYSTRVOYLAELIDCIDLKDYEA--ELCKLIFEMKEEIRHHRMDT	395
Db	901	-----TSTAASSPBY--VAISYVK-----WKSSGDSSTALTSIKRPLPANNVT	943
QY	396	EDLTNRKAGS---NKSLF-----AAANEELIESIRAK--KGDIDNVSPETEKKD	441
Db	944	FSGTPLLRLPALLENPLDLKISNADQKLAAEALISPLTKRKSNOKEKEVKESEQSEKLL	1003
QY	442	NETENDSKD---AEKNNEEEDDLSLEKDSDDKPRDDPQGGSE-----EETEVGD	489
Db	1004	SPYAGTKRSCGAAYOEYEVQPPKPEAPQPTSTTTPPSASNGASHGVPOALAGETMARGD	1063
QY	490	KGNSVSNAGDNTNATSEETSPSEGRPGVCLSE--TPDSSNMAEKKVASELPQDV--	544
Db	1064	TLKQRIDLPGRPKXKHCGRASLIVLGO--PLATGEGDQOEEDADMOQELAAVPSAIWEP	1122
QY	545	-PEEPNKTCESSNTSATTTISIQPLENSNSSSELSNQSSESAKAADDPENGESRESHTVS	603
Db	1123	SPEKPTHTIVTDENDNCASCSTSPVGDSEKSKSGSAQAEVKKA-----TA	1168
QY	604	IQEELVGDFTSEKTEGLESPPGAGKAGASSTHTITRLNPNPSKLSQLKSQVAAAHBA	663
Db	1169	LQKE-----TASABAGSSAKVT--TNN-----AAVAS	1194
QY	664	NKLFKEGVELVNVSGEELISRLSTKKEVIKKGNNYFRLQEGKRYVHNOYSTNSFAL	723
Db	1195	NLUIAASRK-----QRNDI--ATSSSVQOSS-----QYQGRKTKHROQRI--L	1236
QY	724	NKHOHREDHD-----KRRHLLA-----HKFCLT	745
Db	1237	ISIDFMEWYDPAEFCQTFGLIVETVYAOALFCFLGSGTGLDPLIFCAACCCEPYHQYCVQ	1298
QY	746	PAGEFKMNGSVHGS---KVLITSLRLTITQLENNIIPSSFPHNASHRAMI-----	795
Db	1297	DEYNLK-----HGSFEDTTLMGSLLETTVN--ASTGPPSSL--NOLTORLUMLCPRCTVC	1347
QY	796	-----KAYOMCSK-----PREFALATALICEVAPVYMLPIW	827
Db	1348	YTCMNSGSKKQCKQCKKNHSHTCGLSKRYLLGADRPILCVANCLKCKSCSTTKV--	1401
QY	828	REFLG-----HTRL-----	836
Db	1402	SKFQGNLPMCTGCFKLRRKKNFCPIQRCYDNDPDLKMMECGDCGOWHSCFEGSLDEQ	1461
QY	837	HRMNSIEEKEEKKKKEKEKQOBEETMQOATWYKTFPYPKHQYWKOK--GEELRVYTGCGW	895
Db	1462	YNLSLTPLESTIEPFLCKCARNESS-----KIKAEWRQAVAEFKASLYVL	1509
QY	896	SMWSTLTHYRVPFLPONTNVNVRKSLSECTKNMDEMDESDKPKCSRPXKIKLEP---	952
Db	1510	KLLSKSRQACALLKLSPRKVV-----RCTGASSNGKQLDPRKAL	1548
QY	953	--DSEKDEVKSGDAKADONEMDISKITEKKDQVYKEILDSDSKPCKEEPEMEVDMMK	1014
Db	1549	QFSSGSDNGJGSD---GESQNSDDYVEFNDQOQO---QOQRANNMKNP--KVRKL-----PCS	1598
QY	1011	TESHVNCQES--SOVDVNVSEGFHLRTSYKKKTKSSKLDGLLEIRIKQFTLEEKORLEKI	1065

Dd	1599	COOHSHSQSILVDT-----KRIKINGSVLSLEENNYMSQVIO-	1639
Qy	1070	KLEGGIKIGKTSSTNSSKNLSESPVITTKAKEGCOSDSMRQESPANNQD-----PED	11222
Dd	1640	-----SNCDELDIAK-----ELISEQFPWFQNETKACTALEED	1674
Qy	1123	LIOGCS-----QSDSVLRMSDPSTTKKLYPKRVLDDVSINSPEKCKQXSIEND	1175
Dd	1675	MFESCSGCGVEDLDADGAGVASVYNHSTSIQAESRSGVLD-----TPL-----	1717
Qy	1176	IEEKVSDIASRGOEPTKSKTGNDFDFDSDKLSADIGTILCKNK-----PKLI	1225
Dd	1718	--EYVDGSGGCIK-MLDTRMCLFCCKSESGEGLSGBEALILCGHDCWHTTCAMMSAV	1774
Qy	1226	QESDPTIVSSSKSALHSSVKS-----TNDRAAT-----PLSRAMD---	1261
Dd	1775	FEEDID---GSLQWHSAAVARGMIKCTVCGNRGATVGCNVBSRGEBHYHPCARSIDCAF	1830
Qy	1262	-PEKGIQDSSS---NSTLENSDPTVSIQDSSEDMITVONSMSISQEPFRFEQVEYLE	1317
Dd	1831	LTDMSMCPAHAKGNALKANGSSVSYTESNFV-----SRPVYELDRKKR---KILE	1861
Qy	1318	PLKCELVSG-----ESTGNCEDLPAKGTGA---NGKKPSQOKKLEERPV-	1359
Dd	1862	PARVOFHIGSLFVQOLAIYPRFSDSEAVVPINFLCSRLIYSSKEP---KVIYEYIART	1938
Qy	1360	--MKSCDQILF---KNPTDKKNNESEKQO-----RTSTFOINGDKNPK	1402
Dd	1939	TIONSSSTFLTALDVGRNYVTADHTNPNSKEVQOLAMQIARHHTSLARSEFLNGSTDMGSE	1998
Qy	1403	ITLGECLKELISERVYSGAVEPKVNNINKIIP---ENDIK--SLTKESAIRPINDV	1457
Dd	1999	.PPNPNSCVPP-----DQNTHEEPOQADLPEPMLTSPLEKELGJSTHGILLMLLV	2051
Qy	1458	IMEDPNEENSSETKSHLSSDAEGNRDSELPF-----STKESDSTOTT	1504
Dd	2052	V---RLKQGGELKDAIF--EDLPHELBDISMLDILFYLDKDTDLFAISEQSKDQMT	2105
Qy	1505	PSASCPESNSVQVEDMEIETSEVKVYTSPTISEESNLSDFIDENGLPI-KNENVN	1563
Dd	2106	S-----NQNONOQAGGANSVS---ICDDPTRNSMTSLG-NGPAPNSVEDAM	2150
Qy	1564	GESKRKYITEVTWTSTVA-----TESKYIYKXKGDQCVTVSPTNCKASKVITTT	1616
Dd	2151	LSAARN---SSQVOMLTKLMPKLDGNSAMATIKRRKLSK-----NLAEGVETLS	2199
Qy	1617	TTVTKLSTPSTGGSVDLISVKEOSKIVVPTVYDLSLTGGTLVMT--VSKEYSTRD-	1673
Dd	2200	SOQRKKKMAIVAGY-----SRQOSISETSVGCVATTSQSVRSKSFYMSAKRFRKSE	2253
Qy	1674	-----KYKIMFSPRPKTRSGTALPSRYKFTVKSTRKSIFVLPNDLKLARKGGR-	1755
Dd	2254	GREAkakRIMQMG-----VDDSITEPRITISGDONISLTAQFSQVACD	2297
Qy	1736	--EVPYFNYNAKPLDLWIPSPRPPTGITWIRLOTQVYSLAGVSLMLRLMASLRMDM	1763
Dd	2298	RCOCYRYNYDAFOR---HLPPSCSPT-----MSSNETESDVS-----	2330
Qy	1764	AAKVPGGGSGRRTSETSELTITTEIIRKRDVGYGIRFECYCIKICPIGVETPEKPTP	1843
Dd	2331	-----GGGMNNATQISABLSBNLQOLALNAGGLN---CLQ-----SATSF	2369
Qy	1844	PO-----RKGLRSALRPKR-----PETPKOTGVIETVWAEEHEL	1861
Dd	2370	POVOSLSGLSGFQIGQIQLOLQPSQISGDSFSLSQDPDPAQDT-----DDQI	2418
Qy	1862	W--EIRAFAEV-----EKEKAQAEQAKRLEQKPEYIAISTSTPT	1923
Dd	2419	YANSIQSLAANLGGGFTLQAPTYTAPAQPLIAVSTNPJGTQOOFIQIPQTMATATTPPAT	2478
Qy	1924	SSTTSITSPAQKVVAVPISGSVTTTGKRVLT--TKGSPAIYTFQONKNFHOTFATWVK-	1960

Db 722 ADTTOJETVVDKSKKKVLKK--TEKDSF-----ISQSETPPVPER 766
QY 777 NIPSEFHPWASHRANWIKAYOMCSKRPFLAL-AILECAKPV--MLPTMREFLGH 833
Db 767 TKP-----ABSEOKIAEVNAKAKOKEVDNLKREAVAKKIADEKLEAE--- 814
QY 834 TRLHMTSIEREKKEVKKKKKEQEEETMOQATWVKTFFPVHQQVWKOKGEEYRVATYG 893
Db 815 ANIKTAIEVAEAAKKOKEQDLKLETEVYSSKKAAEKLELEKQAI--KKAAL----- 865
QY 894 GWSWIKTHYRVEPKLPCTNNVNYKSLGTYKNNDENK----- 933
Db 866 -----ADAVKOKELNENKLEAAKKSADKLELEESAAKSKVSESVK 912
QY 934 DESDKRCSNPKKIKTEPSEKD-EVKSADAKGAD--NEMDIKTEKKQDVKELL 990
Db 913 GEKKTKRAGEKTYOVESEPTSKTIDTKDVGATEPDEPKKIITKKTEKSSSISOKS 972
QY 991 DSDSDRCK-----EEP-----MEVDDMKTESHV-----NCOES 1020
Db 973 ATDEKVKSKOKEODEPTKPAVSETQWTEADSKKQKEDERKLAELIAKTKOEADEK 1032
QY 1021 SQVD-----VYNVSE-----GFHLRTYKKTSSKLDGL-LERRIK-----OF 1058
Db 1033 SKIDAOEKIKKVEDDAARKEKELNDKLESEIATKKASADKLEBQAKAAEVEA 1092
QY 1059 TLEKORLEKIKLEGKIGKSTSTNSKNLSBPVITK--AKEGCSOSMRQEOSPNA 1115
Db 1093 AKQOKEKEDKLKD-----TEASKKAAPAKKLEKQAOIKKKAAGADAVAKKOELE 1144
QY 1116 NNDQPEDLIQOCSQSDSVLRMSDPSTHTNKLXPKRDVLDDVSIRSPTKCPKONSIED 1175
Db 1145 KNR-----LEANKKSAAGKLKIEESAASKQTEBQAKLDAOTKATAE--KQTKLEKD 1197
QY 1176 IEKVSDLASRQEPKSKTKGNDFTIDSKLASADIGTLCKNNKPLIOESDPTIVS- 1234
Db 1198 -EKSTKESKETVDEKPPKK-----VLKKK-----TEKSDSISIQ 1232
QY 1235 ---SSKSAHSSVPKSTNDR--DATPLRAMDFEGLGCD-----SESNTL 1276
Db 1233 KSTKSTVYESAGPSESEIQKADAROKETDEKOKLEAETIAKKSADBKSLLESKLI 1292
QY 1277 ENSSDTVSIOSSSEEDMIVONSNEISIEQFRTREODVEVLEPLKCELVSGESTGNCEDRL 1336
Db 1293 KKAAEVAEAAKKOKEKDEQKLTDEAASKKAAEKLELEKOSHUK----- 1336
QY 1337 PVKGTANAKKPPSOQKKLEER-----PVNKSQDIKIKLANTDKN-----NENRESE 1383
Db 1337 --KAAEVDAVK--KQELEKORLESEAAKTKADAERKLEBQKKRAEIALIEIOKEQ 1392
QY 1384 KKGQRTSTFOINGKDNKPKIYLGKGC-LKEISRSRVSGNVBPKNVINKIIPENDIKSL 1442
Db 1393 KLAQEOBSRLEDEKKAAEKOKLESETKSKOTEAPKESUDEKPKKVLKKTIEKSD-SSI 1451
QY 1443 TYKESAIRPFIN-GDVIMEDFN-----ERNSEETKSHLSSSDAEGNVRDSELT 1491
Db 1452 SQKSKSAKSTVAAETLESDFNVEKKTVOAKEQSPDESTSATIKRDPAQ-----KTEEL 1506
QY 1492 PSTKESDQOTTPPSACSESNVNOVEDMEIETSF--YKKT--SSPTTSE-----EESN 1543
Db 1507 SKODDDEDEKKTYY-----DGKPPKPEDESEATPPKRRVKKKTQKSDVASASLADAVSK 1559
QY 1544 LSNDFIDENGLPINKNENANGESKRTVTTEVTYMTSTVATESKTYIKVEKGDK----- 1597
Db 1560 LSDDVBEKKKKVKKKT-----EKSDSVISE--TSSVDTIKPPES-VEIPEKKEQOMLHNR 1613
QY 1598 ----QTVASSTENCAK-STVTTTTTTTTVKLSPTSTGSDVLIISVKEQSKTVVTTT----- 1647
Db 1614 FSTDSAVESEPKNAKHDDTEKTTDDMMTRKSSA-----IFSDDQSISSSTSSGRRR 1667
QY 1648 -----VYDSLTTTGT-----LVTSMTYSKEYSTDK-----VXL 1677
Db 1668 RRRGTGASKFASDTLALRGDNVEIEAEILLAEDDTYVWKVNGDADALNSRCHMSHTFFRT 1727

QY 1678 MKFSRPKTRSGTALPSYKRFVTKSKTSIFVLPPNDLKKLAKKGIRE-----VPYEN 1731
Db 1728 LIIDEVPTDSGMEITATGCTESHHTLLKVEELPVDFVYLPKRTSGKGEQVLTISVTIN 1787
QY 1732 YNAKPAOLDWPPSPRPTGITWRYRLQTVKSLAGVSLMLRLMLASLMDMDMA--KVPPG 1790
Db 1788 HPIDISKVYW-LKDGKP-LEINKDYSIDYWG--CSVSLTLR-----RAKEDSGKRYVVD 1839
QY 1791 GGSRT-----ETSEIEITTT-----EIKRRDNGPY-----GIREFYC 1824
Db 1840 GVDCSHTLSIOGKPVLKMSVETKPVITVDKDOFSLIVAYDSNPESFSMTVDGKLEFD 1899
QY 1825 IRKTIQPI-----GVPEPK--EPTPKRKLBSALRPKRPKEIPIKQTPVILIE- 1871
Db 1900 GNSRIDVDDGLKLTIRGVSKTIDAGEYKLNKEFEVAKQEDVKNDTPSAPGDVSVK 1959
QY 1872 -----TWABEELELEIRAFARVEKKA-----QAVEQAKKR-----LEQOK 1911
Db 1960 AESDCHIEMTAPTLEDNGAEVTSYV--IEKESGRKKFHKVATVNGKTSYVVDLEIET 2017
QY 1912 PVIYATSTSP-----TSSTTSTISPAQ--KVWAPISGSYTT----- 1947
Db 2018 PYVIRIAVKNFEGTGEFIEFKPVQTSFPOVPVPEPPTIDVNTSTCSLSMPKPIEDGG 2077
QY 1948 -----GTKVLTTRKVGSPATVTFQOKNMFHQTFATWVKOG--QSN 1985
Db 2078 SPVYGDVYKRENEGWMOKMNGEELVFTSFNVA--LSSGKEYEFKIEACNEAGLKN 2134
QY 1986 SGVVOVQOKVGLIIPS-----STGSOQFTSPQPTATVITPNTSG----- 2028
Db 2135 SNVSSKLLIYEGVLPEIILDMPVKVLNDKYEVTWKSQGEKEFVYQYKSDSSISIVAD 2194
QY 2029 SGGTTSNSQ-----VITG-----POIRPG 2047
Db 2195 IGGPRESAATSKCIIDGLREGIPYVRYAARNQGTGFSEPTIIPVVYLADAPVRLKA 2254
QY 2048 MVIYRPILOOSTL-----GKAIRT-----PVMQPGAPQOVMMQIING 2086
Db 2255 IKPVKIPKGBELRLECHAAGHAPETIYKDKGEIIPTEBNEIYVSGMSALIIHELGA 2314
QY 2087 QPVST-AVSAPNTVSTPOKSLTSATSTNIQSSAOPRPOG--QVKLT----- 2135
Db 2315 EDVGLKVLVENHGTAEBAEAVGIDVAHFNSPSELTEIEEGHDILTEOVSEAV 2374
QY 2136 -----MAQITOL-----TQHGAGNGGLVYV--- 2155
Db 2375 VMYKDKKRLVASDRVOFYAMARKRTLRIKGSTADSGYKCEITDGRSRTGEVIVNQ 2434
QY 2156 -----IOGGOT-----TGOLOLIPGVTVLPGGOOLMOAMPNGTV 2193
Db 2435 EPHIIVGPDAIYKDGELMVLFCETSKPVYKVKFKNGVEIWPONKAKIMENDGRATL 2494
QY 2194 QRELTPLATFTATSTTTTIVSTTAAGTEQOROSKLSPO-----MOYHODKTLPPAOS 2248
Db 2495 EIKNEFDKHDIGAVTASVSEKTSAPAKLVEVAPNLIIPTETRIDGTVHAGNEPFA--- 2551
QY 2249 SVGPAAQOTIOPASRPPOTOPOPAPPEVOTQPEVOTOTTVSSHVSEKOPPHAOS 2308
Db 2552 -----VEFSGFPIPTIHLTNNGTPLKALAVTLEYDOSVVRM--KDVITLDSG 2597
QY 2309 KPOVAASOPQSNVQOQSVVRVQSPQTRIRPSRPSOLSPGOQSOQVOTTSQPIPIPT 2368
Db 2598 TVRVAIAES-PLGOCIKELPIKTIID-----KSPACDL-----QREVEYEDSVFLSMQ 2644
QY 2369 SLQIPSOQPOSOPOVQSSQTLSGQTLNOVSVSPS--RPOLQIQOQOP----- 2417
Db 2645 --PLETNGAP-----LGVYIERKAVYNNRMRPGQYKPKPLTFVADLPCN 2689
QY 2418 -----QVIAVPOLOOQ-----VQVLSQISOVVAOIQAOQSVPOQIKQLPIQIQOSSA 2467
Db 2690 QVYGRITIAVNGVESEPCDITVLLTSSSEPPVS--ESSELFVPKIATLRP--QVIVA 2744

QY 2468 VOTHOIONVTVQAAVQOE-QLORVQOLR-----DQOQKKKQOQIEIN----- 2509
Db 2745 VD-----ETKVTLMEECPSTSLYKVERKKVGDSDMLELANTDRNFKDLSLESEYVQ 2800
QY 2510 -----VNTPSKLLIVELIIOKOVMMKNVAIEHLKOKKSMTPAEREONMIVCNV 2561
Db 2801 VTANGIHAVSSSEETNPVKILIVPGEEMPAS-----KTEKTDPAKSSSEKSS--AEET 2852
QY 2562 MKYILDKIDKEEKOAKKRKRREESVEOKRSKOMATKLSALFLFKHEQLRAELIKRALLD 2621
Db 2853 V-----AEKOVDOQOASESTTEAVEEKKTK-----KVYKKKVAEN 2887
QY 2622 KDOI--RVOEELKRDILKKEKDMOLAQATAVAPCPVPVLPVLPAPAPPPSPPPPPG 2679
Db 2888 KGEETLOVKEKELKKGKAVKQVQDSRRGSLQASS----- 2922
QY 2680 VOHTGLSTPLTVASOKRRKRREESKSSKSKKKMIS-----TTSKETKDT 2727
Db 2923 -----DNESVTTTSEKRSEAELEKNEKSKSAKKSTADLEADKAETEKSETKET 2973

RESULT 9
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Mohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8221122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SOCH>
A:Cross-references: EMBL:X65551
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KiH>

Query Match 3.0%; Score 446; DB 2: Length 2897;
Best Local Similarity 18.7%; Pred. No. 9.7e-09;
Matches 559; Conservative 398; Mismatches 1058; Indels 982; Gaps 140;
QY 378 CKILEEMREELIHRMDITEDLTN---KARGSNKSFLAANEELIESIRAKGDDYNKS 433
Db 49 CKI-----EIHEGAILHNSSSTMPQVNGS-----VIDEPYRLKHGDV--ITI 90
QY 434 PETEEKDNKT-ENDSKDAE---KNREEFDOSLEKSDDKTPDDPQKGS-EEPEYVG 488
Db 91 IDRSFRYENESLONGRKSTEPFRKIREQEPARVRSFSSDPDEKADSKAYSKITEGK 150
QY 489 DKGN--VSANLGDNTNATSEPTSSEGRSPVYGLSETPPSSNMAEKKVASLEQDPE 546
Db 151 VSGNDELDEMLPPVTPPLKRG--APTRKSLV---MHTP---VLKRIKEQOPQSGK 201
QY 547 EPNKCESSNTSATTTSTOPLNENSSSELNSQSESAKADDPENGERESHTEPVIOE 606
Db 202 QESGEIHVEVKAQSLVSPAPSPRKTTPVASDQRRRSCKTA-----PAS--- 246
QY 607 EIVGDPTSEKSTGELSESPGAGKSGSTRITIRLNPDS-----KLSQLKSQVAAA 660
Db 247 -----SSKSQTEVPK-----RGGERVATCLOKRVSISRSOHDILOMICKRRSGA 291
QY 661 HEANKLPEGEVILVNVSGEISRLTKK---EVTMKG---NINNYFKLGOGKRV--Y 712
Db 292 SEAN-----LIAKMAWADVKLGAOKOTQTKVHKGPQRSMNRORRPAATPKRPGEV 343
QY 713 HNOYSTN-----SFALNKHQHRDHDKRR--HLAKHFLCLPAGFEK-----WNGS 755
Db 344 HSGPSTGIANSPCTIILICKAHTEKHYHVARPRVYLNFIISNOKMDFKEDLSIAEMFKTP 403

QY 756 VHGSKVLTISTLRITLTQLENNIPSSF-----FHPNMAASHRANWIKAYOMSK-PR 805
Db 404 YKEQDPLT-STCHIAISNENLLGQFOGDTGCEBLLPTSSFGGVNFFSAONAKOPS 462
QY 806 EFALMALALECAVKKPVMLPIWREFLGH-----RLHRTSIRE----- 845
Db 463 D-----KCSASP---PLRQCIREGNVAKTPRNTKTSLETKTSDETEREKTIV 510
QY 846 -----EKEKYK--KEKKOEELTM-----QOATWVYTPPVKHQVKKOG 884
Db 511 STVNSGRSTEFRRNIQKLPVESKSEETNTEIYECILKRGQKATLQO-----RRG 561
QY 885 EERYRTGYGWSWISKTHVYRFPVPLPGNTVNVYKSLSEGTNNMDENDESDK-----R 939
Db 562 EKEKI-----ERPFET-----YKENILAKEN--DEKKAMKRSRTMGO 597
QY 940 KCS--RSPKIKIETPSE--KDEVGK-----SDAKAGADQNMDSIKTEKKDOVREL 989
Db 598 KCAPNSDLTDLKSLPDTLEMKDTPARQONLLOTODHAKAPKSEKKTITKM----- 646
QY 990 LDDSDKPKCE-EPMEVDDMKTESHVNCQESSQVDVNVSE-----GFHLKTS----- 1037
Db 647 -----PCQSLQP-----EPINTPHHTKQQLKASIGKVGVKELLAVGKFTTSGEHTT 695
QY 1038 -----YKKTKR-----SKLDGLERIKQOTLEKQRLKIKLEGITKIG 1079
Db 696 HREPAQDGKSIPTFPESPQOIIDPARVYGM--KKWPRPKEEAOSLEOL---AGFKELF 750
QY 1080 KRTSNSSKULSSPVITTKKEGQSDSMRQEQSPANNQDPE-DLIOGQSDSSVLRMS 1138
Db 751 QPFGSEEMDEKT---TKICKSPPEPVPTPTSTKQMPKRLKADVEEFLALRLK 807
QY 1139 DPSHTTNKLYPKDRYLDVSI-----RSPETKCPKONSLEN--- 1174
Db 808 TPSAGKAMLTTPKAGDEKDIKAPMGTPIVQKDLGATLPGSKRQLOTPREKQOALEDLAG 867
QY 1175 -----DIEKVSDLASRG-----EPTSKTYGNDEFIDDSKLASADIG 1214
Db 868 FKELFQTPGHTEILVAAGKTKRIPCDSPOSDVDPTSTYKOR-----PKRSIRADVAG 921
QY 1215 TLI-CNNKPLLOQESDPTIVSSSKSALSHPKSTNDRD-----ATPLSRANDPBGKLG 1268
Db 922 ELLACNNLMP-----SAGKAMHTPKPSVGBEKKIITFVGIVQKL----- 961
QY 1269 DSESNSTLSSDPTVSIQDSSSEDMIVONSNESISEOFTREQDVEVLEPLKCELVGES 1328
Db 962 -----DLTENLGSKRRTQPKREAOALEDLGFKELRQTPGHTEEA-----VAAKGT 1009
QY 1329 TG-NCEDRLPVKTEBANGKRPQOQKLEERPYNKCSDOIKLANTDKNNRESEKKGO 1387
Db 1010 TKMPCSSPPESADPTPTSTRQPKTPLEKRDYQKELSAK-----KLQTSGE 1057
QY 1388 RTSTPOINKOKKPKIYLKGECLKELSESRYVSGNV-EPKVANNIKITPENOKISLTYE 1446
Db 1058 THTTKVPGEGEDKSINAFETAKOKIDPAVATGSRHHPKTR--EAOQLEDLAG--KE 1113
QY 1447 SAIRPFINDVIMEDPFNENSSSETKSHLLSSDAGNYSLETLPSTKESDSTQTPPS 1506
Db 1114 LFOPT-----VCTDKPTTHEKTKIACRSQPD-----PYDPTT 1146
QY 1507 ASCPESNSVNOVEDMEIETSEVKVYTS-----PITSEESNLNDPIDENGLPIN 1557
Db 1147 SSKPOSKRSLRKVDVEEELFFALRKRTSPSAGKAMHTPKPAVSSEKNLYA--FW--GTPVO 1201
QY 1558 K---NENNVGSEKRTKVTITEVITMTSTVA-----TESKTVIKVEGDKQOTYS 1602
Db 1202 KDLTEENLGSKRRLQTPREKAOALEDLAGEKELFQTRGHTEESMT---NDKTAIVACKS 1258
QY 1603 STENCAKSTVTTTITVTKLSTPSTGSGVDIISVKEOSKTVTTTITVTSILTTGGTLVYS 1662
Db 1259 SGPDLDDKPNPASKRRLKLSLG--KYGVKEELLAVGKLQVTSSETLHTHTPEFGDGSMA 1316
QY 1663 MTVSKEYSTRDVKVLMKPSRPKRTSGTA-----LPSTYRK--FVTKSTYKS 1706


```

Db 1317 FMESPQOILDSASLIGSKROLRTPKGSVEPEDLAGFIEFQTPSHTKESMTNEKTV 1376
Qy 1707 IIVLPNDL-----KTLARKGIREYFYFNKRALDIMPYSPRPTEITWR 1755
Db 1377 STRASOPDLVDTPTSSKPOPKRSKRADTE-----EFLAKKQTPSAG---- 1420
Qy 1756 YRLQTVKSLAGVSLMR-LLMASLKMDDMAKVPBGSGTETSETELETTETLEIKRVD 1814
Db 1421 KMHHPKPAVGEKDIINFLGTPVQKLDOPNLPG- GSNRLQTRREKQALBELT----- 1474
Qy 1815 GPYGR-----FEYCIRKIC--PIGV-ETPKETPPQKGLASS----- 1852
Db 1475 ---GREFLEFQTECTDNPTADEKTYKKILCKSPQSDPADPTPTMKPRKSLKADVEEF 1531
Qy 1853 -ALRPRP-----ETPKOTGVIIETWVAEELELEIRAF---AEVE----- 1892
Db 1532 LAFRLTPSAGAMHPKA-----VGEK---DINTVGTPEVKLDLGNLPGSK 1579
Qy 1893 -----KEKAQAVEQOAKRLEQOKPTVIATSTSTSTSTISPAQKVMAPISGVY 1946
Db 1580 RRPQTPKEKAKLEDLAGKELFQTPGHTEESMTDKITEVCSKSPQRPVKT----- 1633
Qy 1947 TGTKNVLTTRKVGSPATVTFQONKNFHQTFATWVKQGSNGVVOYOQKVL--GIIPSTG 2004
Db 1634 TSSKORLKISLKG-----VGYKEEVLPGKILQTSG 1664
Qy 2005 TSOQFTSQPRTATVYTRPNTSGSGT-----TNSQYI-----TGQIRGMVITR 2053
Db 1665 KTTQ-----HRETAGDKSIKAKESAKOKMLDANVTGMERP-----RT 1706
Qy 2054 P-----LQOSTLGKALITPTVMVOPGAPQVMTQIIRQGPSTAVSAPNTVSFPGSK 2107
Db 1707 PREEASQLEDLAGKELFQTPGHTEESTDDKTKT-----ACKSPSPS 1751
Qy 2108 LFSANSTSIQSSASOPRRPQGOYKLT--MAQLTQLTGHGNOGLVIVIOGQGTGQ 2165
Db 1752 MDTPTST-----RRRPPTPLGRDIYELSLAKOLTO-----TTH 1786
Qy 2166 LQLIP-----QGYTVLPGQOLM-----QAAMNGTVQ-----RFLF-TPU 2201
Db 1787 TPKVDEDEKGINVREFAKOKLDPAASVTSKROPRTPKGAQPLEDLAGIKELFQTV 1846
Qy 2202 ATTATYASTTTTSTTA-----AGTGE--QROSKLS--POMOVHODKTLPPAOSSSVGA 2253
Db 1847 CTDKPTHEKTKIACRSQRPDPVGPITFKPQSKRSLKADVEESLALRKRTPSVGA 1906
Qy 2254 KAQOTA-----QPSARQPTOPQSPAPQEVQTOPEVQTOQTTSV 2294
Db 1907 MDTPRPAGDEKDKAFMGTPVQKLDLPGNLPGSKRMQTPKE-----KAQLEDLAG 1959
Qy 2295 -----HVPEAOPTHAOSKPOVAASOPQSNVQGSPPRVQSPQTRIRPS----- 2341
Db 1960 KELEFQTPETDKPT--TDEKTKIACKS--PQ-----PVPVTPASTKORPRKRLKADV 2010
Qy 2342 -----TPS-----QLSPGOOSQVQTTQSPPIPIQPHTSU-QLP-SOGQPOSQD- 2382
Db 2011 EEEFLALRKRTPSAGKAMQTPKPAVSDEKININTVETPVQKLDLGNLPGSKRPQTPKE 2070
Qy 2383 -----QVQSTOTLSSGOTLNOVSVPSPRQLOIQPOPYIAYP----- 2423
Db 2071 KAEALEDLVGEKELFQTPGHTEESMTDKITEVCSKSPQSEFRTSRSSKORLKIPLVKY 2130
Qy 2424 QLOQOVQVLSQI--QSOVAYQAOQOSGVPOQIKL--QLPIQOQSAVQTHQINQVAV 2479
Db 2131 DKKEEPLAVSKLRTSGETTOHTPEIGDSKSIKAKESFPKOLDBAA-----SVIG 2182
Qy 2480 QAASVQEOQLQRYQOQLRDOQKAK-----QOOIEINVT-----PSKLLIKVEIIO 2524
Db 2183 SRRQLRTRREKAEALEDLVGEKELFSAPGHTEESMTIDKNTIPOKSPPELTDATSTRK 2242
Qy 2525 -----KOVYKHNNAVIEHLKOKKSMT-----PAEREENORMIYCNQVAKYITLIDKDK 2571

```

```

Db 2243 RCPKTRRKEVKEELSAVERLTQTSQSTHTHKEPASGDEGIKVL----- 2287
Qy 2572 EEKQAKKRRKREBSVQKSKQNA-----TKLSALLFKHKEOIRA----- 2611
Db 2288 --KQAKKKPNVVEEPSRRRRAPRKEKAQPLEDLAGFIELSTSHTOESLTAAGATKI 2345
Qy 2612 -----ETLKRRAILDKDLQIEVQBELKRLDKIKKEXKLMLOAQ----- 2649
Db 2346 PCSPPELEVVDJTASTKRLHFRVYQ-----VOVKEPSAVFTQTSGETTDADKEPAGE 2400
Qy 2650 -----ANAVAPCPVPYVLPAPRAPPSP-----PPQVQHGLSTPPL 2691
Db 2401 DKGIKAKESAKQTPAPASVYSGRRRRPRAPRESAQLEDLAGFKDPAAGHTEESMTDK 2460
Qy 2692 PVASQKRKEEESKSSKSKKMIKSTSEKTKQTKLCICKTPDESKFYIGCDRCQN 2751
Db 2461 TTKIPCKSSPELEDYATSSKRRRRTTAQVKEKE--LAAVGK----- 2501
Qy 2752 WYHRCVGILOSEALELIDEVYVPOCOSTEDAMTVLPLTEKDYEGLRVLRISIAQHKMAM 2811
Db 2502 -----LTQTSGE-----TTHDKPEYGEKGTKA----- 2525
Qy 2812 PLEPYDPN-DAPDYGYATKEP-----DLATMEERQYR-RYKELTFVAD 2856
Db 2526 -FKOPAKRNVDAEDYIGSRROPRAPRKEKAQPLEDLASFOELQTPGHTEELANGAD 2581
RESULT 10
F90073
Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratazu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <CUR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447
Query Match 2.9%; Score 440; DB 2; Length 2271;
Best Local Similarity 18.3%; Pred. No. 1.2e-08;
Matches 442; Conservative 332; Mismatches 1000; Indels 642; Gaps 84;
Qy 411 AANEEILEIRAKKGDIDNVKSPETEKKDKNTEENDSKAKNRREFQDQLEKSDOK 470
Db 90 AASDAPLITSELTQSETVGNQNS--TTLEASTSTADTSVTKN-----SSSVQTSNSDT 141
Qy 471 TPDDDEQKSEEPYVGDKGNSVANLGDNTTNATSE--TSPEGRSPVGCLETP 526
Db 142 VSE-----KSEKVT-----STTNSISNOEKLTISESTSS--KNT 177
Qy 527 DSSNNAEKVVASLPODVPEEPKCTCESSNTSATTTSIQPNLNSNSSSELS----- 579
Db 178 SSSDTRKSVASTSTEQPINTSTQSNANNTSQSTTPSSVNLNKTSTSTPAVKLRTF 237
Qy 580 -----SQSEKAKADDPENGERESHPTYSIOBELY-----GDTFSEKSTG--ELS 622
Db 238 SRLAMSTFASAA-----TTTAAVTANFTIVNKNLQKYVTTSGNAVYDOSTGIVTIT 288
Qy 623 ESPGAKGA--SGSTRITRLRNPDSKLSQKSOQVAAAHANKLFKEKG--EVLVNS 678
Db 289 QDAYSQKGAITLGTIRL-----DSN-----KSFHRSQKYNLGNKKYEGHNGGDGIGFAPS 337
Qy 679 QGEISRLSTKKEVINKGINNYFKLGQEGKYRVVYHNQYSTNSFA----- 722

```

Dd	338	PGVIGETGLNCAANGIGIGLENAIGF----	KLDYIHTYTSKPNBAAKRANMDSVAVAGGA	Fg	393
Oy	723	-----LNKHQHEHDHDKRRHLAHKEFCLTPAC-----	EFKWNOSVHGSKVLITST		766
Dd	394	AFVTTDSYGAVTATYTSSTADN----	AAKLNVQPTNNFTQDDIDINVGD--TKVMV--		444
Oy	767	LRLITIOLENNIPSSFPHPMWASHRANWIKAYOMCKPREPFLALA-----			812
Dd	445	-----KYAGQTMTRNIDMDLAK-----	SGTTFSLMSTASTGATNLQOYOF		486
Oy	813	----ILECAVPVYMLPIW-----	REFIGH-----	TRLHRMTSIRE	845
Dd	487	GTFEYTESAVQVWYVYTTGKDIIEPKTYSGVNDVYTIIDNOQSMLAKGNTYSVDS			546
Oy	846	EKEKVKKKKKOEETMQATWKKYTFPVKH--	OVMKOKEEYR-----		888
Dd	547	YASTYNDNTKVTMTNAGOSVTF--	YFTDVKAPTVAVGNOTIEVCKTNPVTLTTDNGT		604
Oy	889	-----WYGV--GGMSWLSKTHVYRFPKPLBGNINV-----	NYRKSLEGKNNMDENM	933	
Dd	605	GTVNTVYTGPLSPGLSYDSATNSIIGTPKIGOSTVIVSTDOANNKSTTFPINVADTA			664
Oy	934	-----DESDKRCSRPKIKIEP-----	DSEKDEVKGSDAKG	967	
Dd	665	PTVPIPGDQSEVYSPISPIKIAIQTODNSGNAVYNTYTGPLSGLTEDSTNTNITSGTPTIG			724
Oy	968	AD-----	ONEMDISKITEKKDODVKEILDSDSDKPCKEEPM	1004	
Dd	725	TSTISIVSTOASGNKTTTFFKEVYTRNSMDSVSTSGSVOQOSVSTKADS--	QSASTS	782	
Oy	1005	VDDDMKTESHVNCOESSOVUV--	NSEBGFHLRTSYKKTKKSSKLDGLERRIKOFTLEEK	1063	
Dd	783	TSGSIVVSTASSTKSTSVSLSDSVSASKSLSTSESNVSNSVSTSTSLVNSQ-----			833
Oy	1064	QRLKIKLEGKIKIGTSTNSNKNLSSESPVITIKR--	BGQOSDMRQ-----	EDSP	1113
Dd	834	-----SVSSMGSYKSTSLSDSTISNSNTEKSESLSLSTSTDSLRTSLSLSDLSMST			887
Oy	1114	NANDOPEDLIOGCSOSDSSVLBMSPS-----	HTYTKLPKRDVLDDVSI-----	RSPET	1164
Dd	888	SGSLSKQSGLSTSISSGSSSTISASLSOSTNAISTSLSEASSTDSISISNSINOSQA			947
Oy	1165	KCPQONSTIENDIEKVSVDLASRQOEPTKSTKGNDF	IDSKLASADIGITLICKNNKPL	1224	
Dd	948	STKSDSOSTSISLSTSD-----	SKSMSTSESLDSTSTSGSVGSILIASQSV	997	
Oy	1225	IOESDPT-----	VSSSKALMS-----	SYPKSTNRD	1253
Dd	998	STSTSDSKSTSEIYSDSISTSGSLASDSKSMSSVSSMSTSGSTSESLSDOSOSTSD			1057
Oy	1253	ATPLSRAMDEFKLGCCSESNSTLENSDPTVIODSEED-----	MIYONSSESIEDQF	1307	
Dd	1058	SKSLSLS-----	TSQSGSTSTSTSTASVRTSEOSTSGMSASOSDMSISTSFS	1108	
Oy	1308	TREDOVEVLEPLKCEIVSGE-----	STGNCEDRLPYKGTANGKRPQOK	1352	
Dd	1109	DSTSDSKASAPASESISQASASTSGSVSTSLSTSNSE--	RTSPVSDSTSLSTSDSD	1167	
Oy	1353	KLEERPNNKCSDOJK-----	LKTTDDKNNENRESKAKQORTSTP-----	QINGDKNP	1401
Dd	1168	SISES--TSTSDSISEAISAESTSISLSESNSTDSSEOSQASAFLESELSESTSESTSE			1259
Oy	1402	KIYLGECLEKIEISRYVSNVPEKPVNN-----	NKTIIPNDIKSLUTVRESAIRPIN	1454	
Dd	1226	SVSSSTSEESTSLSDSTSESGOSTSTSLSNSTSGASASTSTSTSESTSTPKSSS-----	VS	1280	
Oy	1455	GDVIMEDFNERNESSETKSHLSSSDAEAGNVRDLSLETLPSTKE-----	SDSTQOTYTP--	1505	
Dd	1281	TSLSMNSTSLSNSTSLSTSLSTSDSK--	SDBLSRTSMSTDSOISTSKSDSISTSTSLSG	1339	
Oy	1506	SASCPESNSVNOVEDMEIETS--	EYKKVYTSPTTSEESNLSDNFIDENGILPINKNEN--	1561	

Db	1340	STSESEDSTSSSSKSDSTMSISMNQSTGSTSTSTSLDSTSTLSLSASMQSG	1399
QY	1562	VNGSEKRTVTEVTMT-----STVATSKVIVYVEGDKOTVVSSTENCAKT-----	1611
Db	1400	VDSNMSASASNSTSTSTSEDSQSTSTYVSQSTSQSESTSTSTLSDBSTISKSTSQSG	1459
QY	1612	VTTTTVTXTSTPSTGSGVDIISVKRQKSTVVTVTVDLSLTTT--CGTLVMSVMSKEYS	1670
Db	1460	STSTVASLISGSESDSQSTST--TSASSTYESASTLSDBSTSTNSGASASTSYLSNAS	1518
QY	1671	TRDKVKLMMKFPKPKRTSGALSYKRFKYKTSKRSIFLPLPNDLKLKLARKGIREVPYF	1730
Db	1519	ASE-----SDSSSTSL-----SDSTASAMQSESDSQSTSA-----	1549
QY	1731	NYMAKPALDWPYSPRPFGITWRRKLOTVKSLAGVSLMLRLMLASLRMDMAAKVPPG	1790
Db	1550	-----SLSDSLSTSTSNMSTIASLS-----TSVSTSE-----	1577
QY	1791	GGSTPTESETEITTTLIRRDVGPGLNFCIKKILICPIGPETPKEPTTPARKGR	1850
Db	1578	SGSTSESTSESDSTSTLSDSQSTST-----RSTASQASAST--STSTSDRSTST	1624
QY	1851	SSALPKRPPEPKQTPVILIEFWABEELAMEIRAFAEVEKEXKQAVEQAKRRLBQ	1910
Db	1625	ASTSTSMKST-----SDSQMSLSSTSTSMDSSTLSDS	1666
QY	1911	KPVYASTSTSTPT-----SSTTSTISPAQKVMVAPISG-----	1943
Db	1661	VSDSTSDSTASSTGSMYSLSLSDSTSTSTASLEVMSASISDSQMSSEVNDSESVES	1720
QY	1944	-----SVTTGKMYLTTKKGSATVYTFQOKNMFHOTFATVWKQGSNSGVYQ	1990
Db	1721	NSESDSKMSGSTSVSDSGSLSVSTSLRKSESVSESSLSGSGMSDVSSTSDSSLSVS	1780
QY	1991	VQOKVLIIPESTGTOQTQSTFPQPRATV--TRPMTSGSG-----	2031
Db	1761	TSLSKSESVESDLSDBSKSTSGSTSTSGSLSTSLGSESVSESTSLSDSTISMSDS	1840
QY	2032	-TTSNSOVTTPQJRPQMVTYIRTPLOOSTLGAIIITPVMVOPGAPQVMTQIIRGQVVS	2090
Db	1841	TSTSDDBLSGSLISLSDSTSLSTSDLSLSD--SKSLSSQSM--SGSESTSTSVSDSQSS	1896
QY	2091	TAVSAPMTVSTTPQOKSLTSATSTSNIQSSAS-----QPPRPOGGVYKLTMAQLTQLTG	2145
Db	1897	TSNSQFOSMSTISASESDSMSTSDSSNIGSNSTSTSLSTSDSMGSGSVSTST--TSLSDS	1954
QY	2146	HGNGNGLTVIYQGGQJRTGLOQLIPQVTVLPBPGQLOQAMAMPNCTVGRLEFPPLATTA	2205
Db	1955	ISGS--TSVSDSSSTST-----STLSDSMSQSTSTSTASQSLSTSTSTMSMSA	2003
QY	2206	TTASTTTTTVSTTAAGIGEOROSRLSPOMOVHODKTLPLPAQSSVGPAPAKQAPQAPAR	2265
Db	2004	STSSSQSTSVSTSLIS--TSDLSDBSTSLISIGSGSYSEESTSDSTSLISDBELSTSDSDS	2066
QY	2266	POPOTQPSPAQPEVQOTPEVQOTTVVSHVPSEAOPTHAQSKKPOVAAQSPQGSNVQGO	2325
Db	2063	TSTSTSDSTSGSTSTSIESTSLSTSGSGSTSVSDSTSMSESDSTST--VMSQDKAD----	2115
QY	2326	SPVAKVQSTQRIRPSPOLSPEQOSQOVTTTISOQIPIQPHSLDIPBQGGPQSQPOVQ	2385
Db	2116	-----STSLISDBESVSTSTSTSLSTSDSTSTSESLSTSMGSGQIS	2156
QY	2386	SSOTLSSGOTLNGVSVSPSPROLOIQOQPOQVIAVPLQOQOVV--LSQISQVAVQIO	2444
Db	2157	DSTSTMSMGST--STSESNMHPBDSKMHHTHTSTSTSRLSSEATTTSTESQSTLSATSE	2214
QY	2445	-AQQSGVPOOLIKLOLP 2459	
Db	2215	VTKHNGTPAQSEKRLP 2250	
RESULT	11		
	A48666		

cell proliferation antigen Ki-67, long form - human

C.Species: Homo sapiens (man)

C.Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999

C.Accession: A48666

R.Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde

J. Cell Biol. 123, 513-522, 1993

A.Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq

ins.

A.Reference number: A48666; MUID:94043435; PMID:8227122

A.Accession: A48666

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-3256 <SCH>

A.Cross-References: EMBL:X65550; NID:9415818; PIDN:CAA6519.1; PID:9415819

C:Superfamily: kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 439.5; DB 2; Length 3256;

Best Local Similarity 18.3%; Pred. No. 2e-08;

Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;

```

QY 300 VPGVTGVAIEIOKNKRYIRHEPIGYDRSRKRYFLNRRLIIEEDTENENKKTWY----- 354
DB 221 VP--TTCQLDNKSKNE-----SPEWKLYESVKKELDVAKSOKEENVLYQYRKSG 265
QY 355 ---YSTQVLAELIDCLDKDWEALCKILEEMREEIHRHMDITPDLNFKARSGNKSFL 410
DB 266 LQDVTATEKESADGLO-----GETQLL-----VSRKSRKSGSGSHANA 304
QY 411 AAN--EELIESTRAKKGIDNVKSPET-----EKQKNET-----ENDSKDAEKNR 455
DB 305 EPASPEQLELDONKGRDYESVOTPSKAVGASEPLTEPAKMKTPYVSOQONSPQKHKK 364
QY 456 EEP-----EDOSLEKD-----SDDKTPDDDEQCKSEEPVEVQKNGSVS--ANLGNTN 504
DB 365 DLYTTTRRSVNLGKSEGRKAGDKITLTPRKLTSTRNTPAKVEDADATKPEVLSKTRG 424
QY 505 A--TSEETSPSE--GRSPVGLSET-----PDSSNAEK-----KVASELPDQVPE 546
DB 425 SIPTDEVLPTEIENEFELTMLQVERKIQKSLKPEKIGTGAQMGCSLPLDSSV 484
QY 547 EPRKCESSNTS-----ATTSTIOPNL--ENSNSSELSNQSOSAKAA----- 588
DB 485 DINNFQDSINSEGLPLKRRRVSEFGHLRPELFEDENLPENTPLKREAPTKRSLVMHP 544
QY 589 -----DDPE--NGERES-----HPPVSTIOEITVG-----DPTSKSGELSESPG 626
DB 545 PVUKTKIKQPOPQSGKQSGSEIHVEVKAQSLVISPAPSPKTPVASDQRRRSCKTAPA 604
QY 627 AGKGAS-----GSTRILITRLRNPD-----KLSQLKSQOYAAAAHEANKLFEKGEVL 674
DB 605 SSSKSTQTEVPRKGRGERVATCLOKRVASISRSQHDILQMICSKRRSGASEAN-----LI 656
QY 675 VYNSGGEISRLSTKK--EVIMKG--NINNTFKLGQBEKRYV--YHNOYSTN-----S 720
DB 657 VAKSMADVVKLAKQOTQTVIKHGPQSMNKKQRPAPRPKPKVEGVHSGFSTGHANSPCT 716
QY 721 FALNKHQHREDHDKRR--HIAHKFCLTPAGEFK-----WNGSVHSGKVLITLITL 769
DB 717 IIGKAHTEKVAVPAPRYVANNFISNOKMDFKEDLSGIAEMFKTVKRPOLT--STCHI 775
QY 770 TITTOLENNITPSSF-----FHPNMAISHRANWIKAVOMCSK--PREFALALALIECAV 819
DB 776 AISENENLGNQFOGDSGEPELPTSESFQGNVFFSAONAKOPSDD-----KCSAS 827
QY 820 PYVMLPIWPEFLGH-------RLHRMSTIERE-----EKKK 849
DB 828 P-----PLRRQCIIRNGNNAKTPRNTYKMTSLTKTSDTFTEPSKTVSVNRSGRSTEF 883
QY 850 VKR--KEKKQEEETM-----QOATWVYKTEPVKQVVKQKQGEVRYVGYGWSMI 898
DB 884 ICKLPEVESSEETNEIYECILKRGOKATLLQO-----RRGEMKET----- 925

```

```

QY 899 SKTHYRVFVPKLEGNTNVNRYKSLLEGTKNNMDENMESDK-----RKCS--RSPKKIKIE 951
DB 926 -----ERPET--YKENIELKEN--DEKRAMKRSRTWGOKACAMSDTLDKSL 970
QY 952 PDSE--KDEYVG-----SDPAKAQONEMDISKITEKKQDVQKELLDSDSPCKE--EP 1002
DB 971 PDEELMKDVARQGNLQTDQDHAKAPKSEKGI TKM-----PCOSLOP 1012
QY 1003 MEYDDMKTESHVNCQESSQVDVNVSE-----GFHLRTS-----Y 1038
DB 1013 ---EPINPTHTKQOLKASIGKVYKEILLAVGKFTPTSGETTHHRRPAGDGKSI 1068
QY 1039 KKKTK-----SSKLDLLEIRIQFLPEKQREKIKLEGIGIGIKTSTNSKNLSEEP 1093
DB 1069 KESPQIILPRAVVGDM--KKMPRTKEAQSLIEDL--AGFELFLQTPGSPSESM 1123
QY 1094 VITKAKGCGOSDMROESQPNANNDOPE--DLIOGCSQSSSVLRMSDPHTTNKLYPKDR 1152
DB 1124 T---TIAKSPPEPVDPDTPTSTKQMPKRSIRKADVEEFIALRKILTPSAGKAMLLPKPA 1180
QY 1153 VLDDVSI-----RSPETKCPKQNSIEN-----DIEEKVSD 1183
DB 1181 GGEKDIKAFMGTPVQKLDLAGTLPGSKROLQTPKEKAQALDLAGFKELFQTPGHTBEL 1240
QY 1184 ASRGO-----EPTSKTKGNDFFIDSKLASADDIGTL--CKNKKPLIQE 1227
DB 1241 VAAGKTTKIPQDSPOSDPVDPTPTSTKQ--PRSTRKADVEBELACRLMP----- 1290
QY 1228 ESDTISSSKSALHSSVPKSTNDRO-----APPLSRAMPDEKLGDSSESTLENSSDT 1282
DB 1291 ---SAGKAMHTPKPVGEEKDIIIFVGPVQKL-----DLTEVLGTS 1329
QY 1283 VSIQDSSEEDMLVQNSNESISQGFRTREQDVLEPLKELVSGESTG--NCDRLPVKGT 1341
DB 1330 KRRPQTPKEEAQALDLDLTFGKELFQTPGHTEDA-----VAAGKTKMKPCSSPESAD 1382
QY 1342 EANGKRPQOQKLEERPVRVKSCDOIKLKNTDKNKNNESESEKQORTPTQINGDNRP 1401
DB 1383 TPTSTRQPKTLEKRDYOKELSAK-----KLTQSGTTHTDKVPGEDEKS 1430
QY 1402 KTYLGECEKLEISESRVSGNV--EPRVNNINKIIPENDIKSLTVKESAIRPINDVIME 1460
DB 1431 INAFRETAQOKLDPASVYSGRNHPKTK--EKAQPLELAGN--KELFQTP-----VCT 1480
QY 1461 DENERSSETKHILSSPABGNYSLETLPTSKESDQTQTTSPASQSPESNSVQYVD 1520
DB 1481 DKPTTHEKTTKTAQRSQD-----PVDTPSSKPOSKSLRYVD 1519
QY 1521 MEIETSEVKKVYS-----PTSEESNLNDFFIDENGLPIPK--NENVNGESKR 1568
DB 1520 VEEEFALAKRTPPSAGKAMHTPKPAVSGEKNIYA--FM---GTPOKDLITENLNGSKRR 1574
QY 1569 KTVIIEVTMTSTVA-----TESKTVIYKQCDQTPVSVSENCAKSTVTTT 1616
DB 1575 IQTPKEKAQALDLAGFKELFQTRGHTESMT--NDKTAKAACXSQSDPDLKNPASSR 1631
QY 1617 TTVTKLSTPSTGSDIISVKQESKTVVTTVTDSLTGGLVMSMYKSEYSTRDKV 1676
DB 1632 RLKTSIG--KVGKELLAVGLTQTSGETTYTHHPIDGSKMAFMSPEKQIILDSN 1689
QY 1716 -----KKLARGGIREVYENYNAKPAALDIPYSPRPTGKITRYKLOTVKSLAGVSL 1769
DB 1750 SSKPOPKRSLRADTEE-----EFLAFKQTPSAG--KAMHTPKPAVGEK 1793
QY 1770 MLR--LIMASLRMDMAAKVPPGSGSTRTSETTEITTEIIRRDVGPYGR----- 1820
DB 1794 DINTLGLTPVQKLDQPGNLP--GSNRRLOTRKEKAQALBELT-----GREFLEFQTC 1844

```


F.129-161/Domain	ankyrin	repeat	homology	<AN03>
F.162-190/Domain	ankyrin	repeat	homology	<AN04>
F.191-223/Domain	ankyrin	repeat	homology	<AN05>
F.232-264/Domain	ankyrin	repeat	homology	<AN06>
F.265-287/Domain	ankyrin	repeat	homology	<AN07>
F.298-330/Domain	ankyrin	repeat	homology	<AN08>
F.331-363/Domain	ankyrin	repeat	homology	<AN09>
F.364-396/Domain	ankyrin	repeat	homology	<AN10>
F.397-429/Domain	ankyrin	repeat	homology	<AN11>
F.430-462/Domain	ankyrin	repeat	homology	<AN12>
F.463-495/Domain	ankyrin	repeat	homology	<AN13>
F.496-528/Domain	ankyrin	repeat	homology	<AN14>
F.529-561/Domain	ankyrin	repeat	homology	<AN15>
F.562-594/Domain	ankyrin	repeat	homology	<AN16>
F.595-627/Domain	ankyrin	repeat	homology	<AN17>
F.628-660/Domain	ankyrin	repeat	homology	<AN18>
F.661-693/Domain	ankyrin	repeat	homology	<AN19>
F.694-726/Domain	ankyrin	repeat	homology	<AN20>
F.727-759/Domain	ankyrin	repeat	homology	<AN21>
F.760-792/Domain	ankyrin	repeat	homology	<AN22>
F.793-825/Domain	ankyrin	repeat	homology	<AN23>

```
QY 1800 EREITTTTIRKRDVGPVIGIREYCIRKILICIPV-----ETPKETPTPORKGLRS 1851
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2888 DPQITS-----PYE-----NWPSQSFSSSESKTQTDANHTTSPHS 2923
QY 1852 SALRKRRREP-----KOTGPVILETWVAEELMEIFAFAERVAKERKAQAEQA 1903
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2924 SEVSVVTTSPEDVAVVASSSSGTALSK-----ESNFEQDOK-----MESOL 2966
QY 1904 KRL-EOKRPVIATSTSTSTSTISPAOKVMVAPISGVTTGKMWL-TKVGSA 1961
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2967 ESTLWMOQSDSV--SSSEPTMSATTTVGEQI-----SKVITTKTDVDS- 3010
QY 1962 IVTFQOKNKHQTFATWVQOGSNGVVOVQKVLGILIPSSGTSGQTFSPQPRATVY 2021
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3011 --SWEIRREDDAEFARVY-----EEQKIFGLM-----VDRQ----- 3041
QY 2022 IPRNTPSGSGTSSNGOVTIPQIRPGMVIIRTPLOOSTIGKALIRPVWQCAPQOVMT 2081
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3042 -----SQGT-----PDTTPARPTPEGT--PTSEQNPLFOEGK---LF 3076
QY 2082 OIIRGQPVSTAVSAPMTVS---STPGOKSLTSTSTSNIOSSASQPPRQOGVKLTMAQ 2138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3077 EMTRGSAIMTKRSTADESFHFQIGQESRETLSEDVKEGATGADPLLE-----TSAE 3131
QY 2139 LPQLTQGHGNOGLVTVVIOGOGOTTGOLILPOGYT---VLPFGQOLMOAMPNGTVQ 2194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3132 SLALS-----ESKETVDEADLLPDSVEVEEIPASDAQLN----- 3168
QY 2195 RLFLPPLATTATTATTTTSTTTAACTGEOROSKLSPOWQYHQDKTLPPAOSSSVGPAK 2254
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3169 -----SOMGISASTERTPTKEAVSVGT-----KDLPTVOTGDIPLPS 3204
QY 2255 AQPOTAP--SARPPQOTO-----PQSPAQP-----VQOTQPEVQOTQ 2290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3205 GYKQJSCPPSSSEPAVQVQDLDFSTLRSVYSDKGDSDPSSEPOKSVLEIPTAPMENVP 3264
QY 2291 TVS-SHVPSEAPOT-----HAQSSKP-----QVA 2313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3265 TESKSKIPVTRMPTSPAPPAEYESSVEDFLSSVDEENKADKAPKLPKVPVLQWY 3324
QY 2314 AOSQPOSANVQGSPPAVQSPSOTRIRPSTPSQJSPQOSOVOTTISOPPIPIOPHTSLQIP 2373
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3325 ECOLSDLDTSVQKTVAPQODMASIAPDNRSK--SESDASSLDSKTCPPVKTRSTETETE 3383
QY 2374 SOGQOP-----SOPVOYSTQTLSSGOTLNOV--SVSSPSRQOL----- 2411
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3384 SKEBAELESEEGATRKILITSLPVKSRSTTSSCRGCTSTPTKSKHFFDLYNSTE 3443
QY 2412 -----IQOPQOVIAVPOLOQOVVLSQIQOVVAVQIQAOQSGVPOIKLOLPIQIQSS 2466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3444 PFEETIDEASKLVDRLTQSEREGEIVSDESSALEVSIENLPRVETSHSVEDLFFDR 3503
QY 2467 AVOTHOIONVVT-----VOAASVOEOLORYO-----OLRDOQOKKO 2503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3504 PWDSEIETLRIPIEDENGHDAEPDQDEQRIEERLAYIADHLGFSWTELARELFTEE 3563
QY 2504 QOIEINVTNPKL-----LKV-----EIOKOVYMKHNAVIEHL 2538
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3564 QIHQRIENPNLSLODOSYLILKIMLERDGHATDTNLVLECLTKINMDVHLMETPEPL 3623
QY 2539 KOKKSMTPAREENOMIYCNQVMKYILD-----KIDKEEKOAAKKRKESEVEOKRS-- 2591
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3624 QBRISHSVAFIEQIOT-----ITLDHSEGSFVLOJELCTAGHOKKQKEQAVSSESET 3672
QY 2592 -----KONATKLSALLF--KHKQLRREILIKRRALLDKDIQ 2625
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3673 CDHPRIVSEEDISVGSYTFQDGVNPTGEGSSSTALFPQTHKEQVOODFGSKMODLPEESS 3732
QY 2626 IYVQELKMDKLIKKEKMDQLAQATAVAAPCPVTP-----VLRPARPARPPSPRPPEV 2680
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3733 LEVQOY-----FVTPPTGETSETOKAMIVPSSKPTPEVSTPAE 3773
QY 2681 QHTGLSTPT-----LPVASOKRKRKEEKSSSKKKMKMISTSTSKTKKDTKLYCICK 2734
```

```
Db 3774 EKKLYQTPSTSESGSPPIIOEPEPSEHREES--PRKTSLVIVESADQPEL-----CE 3827
QY 2735 TPYDESKFYGICDRCONWVHRCVGLIOSEAEELIDYVPOCOSTFDAMTVLPLTEK-- 2792
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3828 RLDEDAFEKGDDMP-----IPPEVTEEYI-----DEHGHTVVKKTVRKII 3871
QY 2793 ---DYGLKRVLRSLQAHKMAFPLEPYDPNDAPDPYGVIR 2830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3872 RRVYSEGTKEEIMWQMPQ-----EPVNIERGEQYSKVIK 3908
```

RESULT 13

```
T30249
cell proliferation antigen ki-67 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence: revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30249
J:Starborg, M.; Gell, K.; Brundell, E.; Hoog, C.
J. Cell Sci. 109, 143-153, 1996
A>Title: The murine Ki-67 cell proliferation antigen accumulates in the nucleolar and
1 for cell cycle progression.
A:Reference number: Z20787; MCID:96431717; PMID:8834799
A:Accession: T30249
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2938 <SWA>
A:Cross-references: EMBL:X82786; NID:q1177527; PIDN:CA58026.1; PID:q1177528
A:Experimental source: strain CBA; testis
C:Genetics:
A:Gene: Ki-67
C:Keywords: cell cycle control; nucleus; tandem repeat
```

Query Match 2.9%; Score 439; DB 2; Length 2938;
Best Local Similarity 18.0%; Pred. No. 1.8e-08;
Matches 537; Conservative 461; Mismatches 1079; Indels 914; Gaps 135;

```
QY 335 NRRLIEEDTENENEKKI---WYSTKVQLAELICLDKDYAEALCKILEEEREIHR 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 NSRLTQSLSNSNEKESPFKLYQSMKEEL-----DYVSQKSCRSKSEPPQPRAAE 260
QY 391 HMDITEDLTN--KANGSKSFLAANEELIESIRAK--KGIDNVKSPETEKDNET- 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 ESRETOLIVSGRARAKSSSTPYTAASSPKVGIWTERNRGAVPVQSTETAKMKTPIVR 320
QY 445 -----ENSKDAEKREBEFDQSLKDDDKTPDDPQKSGSEETVGDGKNSANLU 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 HSOQLKDEDSRYVGRHRSVNLDEGSAQAVHKTVTPGKLATRQOTPEAGDVGSP----- 375
QY 499 GDMTMTATSEETS--PSEGRSPVQCLSETPPDSSNMAEKKVASE--LRQVDEEPNKTCESS 555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 ADTPEHSSPQNSIPAKVNP--SAETONRLSLTORLVGEEKTKGSGFSKREKATATA 432
QY 556 NTSATTTSIQPNLENSNSSSELNQSQSESAKAA-----DD--PENGRESH 599
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 EGTCSGLPLGSSVDISNFGDSINKSEGMKRRRVAFGHLRLPELDENLPRVTPLRGE 492
QY 600 TPVSIQELVIGFTSEKSGELS-----ESPG-----ACK 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 TP-----TKRKSLGTHSPAVLAKTIKERPOSPGKQESPGLTPRTINDQRRRSGR 541
QY 630 GASGS-----TRITTLRN-----PDSKLSOLKSQ-----QVAAAHANEANKLFKG 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 TSSGSNFLCETDIPKAGAKSGMLPKRASISRSQHGIIQIMCSKRSASAEAN----- 595
QY 671 KEVLVNSQGEISRLSTK-----KEVIMK-----GNINNYFKLG-- 704
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 --LIYAKSWADYKLGQVQTKVAKHVPKQTSKQRBSPRPKRPKTSMLHNOFTTGIAN 653
QY 705 -----QEGKTVYVINOYS--TNSFALNKHQHEHDDKRRHLAKKCLPRPAGFFKAN 753
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 654 SPCTIVVAGRAQIEKVSVPAPRYKMLNNMLNLR---KVDESEDLISGLTEMFKTPVKEKQOO 710
```



```

Db 2604 --SRKRLRTHKNGMSITLLKLLGSGKEITQ--ISDSHSEKLAHDITSLKSTQOQK--DS 26255
Oy 2552 NQRIIVGNQYWK-----YILDKID-----KEEQQAKKRK--REESYEQKRS--- 2591
Db 2656 VKPLRTRRRLVLRASKVEKPEVLVDTRDHTLOSQSNPLSPKRRSARGPSIYTRRLSL 27155
Oy 2592 --KQNAKRLSALLFKHKEQJRAEILKRLALDQDOI-----EVOEELK-----RDLKI 2638
Db 2716 APKQEAETDEKRVPEKKRAASSKRSVSPPEPKMHLLIYSNNKLESVEEOVSIVYMKTEMEA 27755
Oy 2639 KKEKDLMLQQAATAVAAPCPPTV-----VLPAAPAPPSPPPPGVONTGL--LSTPLP 26932
Db 2776 KREN-----PYTPPDQNSRYRKRTKNVQGRPFQASAEWGIKKNEKTMK 2819
Oy 2693 VASQRRKREESQSSSKSKKKKMIIS--TTSKEKADTKILYCICKPIPYDESK 2741
Db 2820 TASQETELQNDQGAKKSTRGVSSSKRCLNSRGTEGTEPPQCEAEKERTSK 2870

```

RESULT 14
T34434
hypothetical protein K06A9.1a - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Gelsel, C., Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-2232 <GET>
A:Cross-references: EMBL:U08046; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

[illegible]

Dd	1084	SS---	TSGGSSTSPNPQOSTISPTSCATSTSPGSGTTLTSLSPBPSOBSSTISSOGSCT	1140
Qy	1528	VKVTYSPITTEEBESNLNPFIDENGLPIKKNBNVNG-----	ESKRKYIT	1573
Dd	1141	VWTTTSGMNTSQSGTOJPG----	STGSTVJPQBTSGSGSTSGEILTSGSGTYOTPRSSLST	1196
Qy	1574	EVTWTST--VATES--KTIVIKVEKDQOTVSVSTENCACKSTVTTTT-	-----TW	1619
Dd	1197	SPAISTSTQQGVSNPNSGSHVTO-----	PSTVGSTYS--SGSTVTYGSTEGSSTSGSSA	1249
Qy	1620	TKLST-----	PSYGSVDIISYEKGSKTVVTTTTVDSL- -----TTGGT	1658
Dd	1250	TSLSASSPVSTSQSPNPBSTSGSSTPTPNPDSQTSVPVSTTTGEMMSHGSTQTPSTIGST	1309	
Qy	1659	LVTSMKVSEKEYSTRDKYLIMKFSPKKTRSGTLLPSYRKFVKTKRSSTIFVLPNDLKL	1718	
Dd	1310	VYQPBTVSGSNSSSYTYI---GSSEASTSGSSF--	KTBPSLSIPVPTS--	1354
Qy	1719	ARKGIREVPYFNYNAPALDIMPYPSPRPTF-GITWRYRLDTYKSLAGVSLMLRLMAS	1777	
Dd	1355	-----	PIPTTFASSTGSTISDVSVSTTSLA -----	1382
Qy	1778	LRMDMAKVPVPGGSGRTSETSELEITTHILIRRDVGYXGRFEJCIRKIICPLGET	1837	
Dd	1383	-----PLSSLP-----	STVPSTQSFSSTSGSSKASSP--	VSQ 1417
Qy	1838	PKERTPORKGLRSSALRPK-----	RPEPKQGTGPVILETWABEE	1878
Dd	1418	TSSPTTNPTGTBESTLISLTISGTHOTTMKASAGSSTSPSTNSQTGTVMG----	1471	
Qy	1879	LEMEIRAFAERVEKEKAQAEQOAKKRLBOOKPTYIATSTTSPSTSTSIIPAOKVNV	1938	
Dd	1472	-----SSSTSGVSTSSASSPTQMSTSQOSSAGSTVVASSTPAASSAPBSTG-TMS	1523	
Qy	1939	APISSVTTGTKMYLTKRVSP--	ATVTFQONKNHQTFFAMWKAGOSNCSVVOYQKV	1995
Dd	1524	STSGCVSTSIIEBSTTASASQOTGTVTMGSSST-----	SGVSTSSASTOPOM	1573
Qy	1996	LGIIIPSSGTGSOQTFEFQORFATVTLRPTMTSGGTTNSOVINGPQRPMGTVIRPPL	2055	
Dd	1574	STSGSSAGSTVVASSTIGLVSTVST--	BSTGTMGSTSG-----TVGSTIS	1618
Qy	2056	QOSTLGAIIITPYMVOPGAPPOQVMQOLINGQGVSTAANAPTAVSTPQO-----	KS-L	2108
Dd	1619	ESSTTTAASSQOTGTVTMGSS-----	ISGVSTSSASTOPQMSTSQOSSAGSTVAST	1672
Qy	2109	TSASTNSTIOSSASQPPRPQGOGYKLTMAOLTLOGHGNOGLTVIYOQOQT-----	2162	
Dd	1673	TGLVSTSTVPBSTGMSTSGSTGVTISESS--	TAAASQOTGTVTMGSSSTSGVSTS	1730
Qy	2163	---NGOLOLIPO-----	GUVVLEGPQQLQOAMP-----NGVQRFLEFPPLATT	2204
Dd	1731	SASSQOPQMSTSQSSAGSTVIVSTVASPAASSTAPSTGMTSSTSGTVG--	STMOS	1786
Qy	2205	ATTASTTTTYYSTTAAGTGEOROSKLSPOMQVHQD-----	2240	
Dd	1787	STATSTHTGOSTYTLLS-----	STSSNMOSTSQGSVGSVASTAGLIVSTVPSSTG	1842
Qy	2241	TLPPAOSSVG-----	PAKOPOTAQ----PSARPOLOTPOPSAPOPEVOTQPEVO	2287
Dd	1843	TMGSTSGCVSTSIIESSJTASASQOTGTVTMGSSSTGCVSTSSASSTOPQMSTSQCSS	1902	
Qy	2288	TQTVYSSH-----	VPBEAQPIHAOS-----SKPOVAASQOPQSNVNGGSPRYR	2330
Dd	1903	AGSTVASTTAGLVSTVPBSTGTMGSTSGTVGISTSIESTTAASTSTSGTC--	STVYTI	1958
Qy	2331	QSPQOTRIPTSPOLSPGOOSQOVOTTTSORPIQPHNTSLQIPSGOPOPQPOVQOYST-Q	2389	
Dd	1959	GSTGCT--NPESPMSL--	SQI-TITTEPSSQSTSTQTSILPSSPSPSTHSVSSBECT	2010
Qy	2390	TLSSQOTLN-----	QVSVSPSRPOLIQOQPQOVIAVPOLOQOVYLSQ	2434

Db 2011 TMSGATTSDKMSFLSSTGTYFSRSGSSLATTSAKPSTVCLMYTPQSKIEDO 2067

RESULT 15

D71623

erythrocyte membrane protein PfEMP3 PFB0095c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: D71623

R:Cardier, M.J.; Tecteljin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MID:99021743; PMID:9804551

A:Accession: D71623

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2441 <GAP>

A:Cross-references: GB:AE001371; GB:AE001362; MID:93845092; PIDN:AACT1809.1; PID:9384509

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0095c

Query Match 2.8%; Score 417; DB 2; Length 2441;

Best Local Similarity 18.6%; Pred. No. 9.2e-08;

Matches 525; Conservative 424; Mismatches 1053; Indels 824; Gaps 132;

QY 215 YGFVENKIVLQPLVDOFLTNTARELMSGY-----IQYDHCRCYHKLGLD 263

Db 44 YNKIDVNVNFEILRLKSLAQVGNRLSRGVRDPRTEALKKQFROHKK----- 95

QY 264 LCCETSAVYHLECYAPPELEVPEDMOCEVCY-----AHKVPGYDCAEIQ 311

Db 96 -----KEALKQKTEKNEKARNALKEKKLEQKKNDQAKADLKKESSQS 140

QY 312 KKKPYIRHEPIDGYDRRRKWFNLRLIEEDTENENKKIWTYSKYQ----- 360

Db 141 SSKSLAKEVNGALKEK-----NKETLKKLELNOKEKE--EKKKIDONDEALKNG 193

QY 361 -----LAELTDLDKDYEAELCKILEEMREELHRMDITTELTNARGSKSFLAA 412

Db 194 NDKDKKIVKPKPESEYKDLKEWL-----KEKEFIKQLKDYEEKKKRR----- 239

QY 413 ANELLESTRAKGIDTNKSPETEKDNETP-----NDSKDAKKNEER--EDQSEKUS 467

Db 240 --NWILRSILRR-----DKREIQLKLEKLAOESALNEIKERRASRRPMVVMQRMKDE 292

QY 468 DD---KTPDDDEPGKSEEPTEYVGKGS-----VSANLGDNTTNAETSETPSEGRSPVG 520

Db 293 VDEMICKYDDEQAEKNGTDEELIKDGDYEEILVETKFTYGKRENA----- 337

QY 521 CUSEPDDSSNMAEK-----VASELPDVDEEPPKTCESSTSAATTSIQPNLENSNS 573

Db 338 -LGEIDEEEREYKRRKRYLKEDGEDL-KDVEEKLREETGYGFRKPTTRILVKKRRNKE 395

QY 574 SSELNSSQSESKAADDPENGRESITPVSIOELVYGDTSKESIGELSESFGAGKASG 633

Db 396 QKKLEDEKKEKKLIAAEPD-EKKIKLKD-DKKVYVNVKNNKS----- 438

QY 634 STRILIRLNPDSKLSQLSQVAAAHBANKLF-----KEGEVLYVNSQGEISRLSTK 689

Db 439 ---FPDKFAPDKKRTM-----YLSLELFTVPPKDN-LAVSDCMDSKNGKK 485

QY 690 EYIMKGINNYFKLOGEKYRYVHNOYSTNSPALN--KHQHREDHDKRHNLAKFCLTPA 747

Db 486 ---LSTFNPFRKRRNKLKER--KMQELHKFKKVKYQKLEBEKKRN-----PD 531

QY 748 GEFKNGSVHSGKVLITSLRLITQLENNITSSFFHPWASHRAMKIAVOMCSKPRFF 807

Db 532 GELPLTPEIH---VLRPSD-----MDGENKKSAGHH-FKYQPTPKKEYEESHVADY 581

QY 808 AALALILECAVPRVVMPLWREFLGHTRLHMTSIEREEKVKKKKKKQEEELMQAAT 867

Db 867

Db 582 QL-----EHEPPTKLPEYK--GH--VSREYQLDNEVADLEPEYKGVHRE----- 624

QY 868 WKTTFPVKHQVKKQGEERYATGYGAWMSKTXHYRRFPKLPQNTNVRYSRLBGTK- 926

Db 625 ---YOLDNEV-RDELPEY-----EGHISR-----EYOLDNEGPRST 656

QY 927 -----NNDENDESDKRCSSPKIKIEPDSEKDEYKGSDAK--- 967

Db 657 LKEYDQTELAQKIDITNKPHEVDEYDQSELAKG-KDITNKPHEVDEYDQTELAQKKEY 715

QY 968 -----ADQNMDSK---ITEKKDQYKELLSD--SDRKCKEPEHE-VDDDMKTE- 1012

Db 716 TNKPHENLEEYNETDLAKGEYTNKPHEVDEYDQSELAKGKIDITNKPHEVDEYDQTEL 775

QY 1013 -----SHVNCQSSQDVNVNVEGPHLRTSYKTKKTSKLDLRLR-----RK 1056

Db 776 AKCKEYTNKPHENLEEYNETDLA-----KGKEYTNKAHNELEEYNETDLAKG 823

QY 1057 QFTLEEKORLEKIKLBGIGIKGTSTNNS-KNLSE-----SPVITKAKEGC--- 1102

Db 824 EYTNKAHNELEEYNETDLAK--GKEYTNKAHNELEEYNETDLAKGEYTNKAHNELEEYN 881

QY 1103 QSDSMRQEOSPANNQDPE-----DLIOCSOSDSVSLKMSPSHT-----TNKLYP 1149

Db 882 EYDLAKGEYTNKARENLEEYNETDLAKGEYTNKAHNELEEYNETDLAKGEYTNKAHE 941

QY 1150 KDRVLDDVSIKSPETKCPKONSTIENDIEEKY-----SDLASRGOEPTKSKTGPNPFID 1204

Db 942 N---LEEYN---EYDLAKGEYTNKAHNELEEYNETDLA-KGEYTNKAHNELEEYNE 992

QY 1205 SKLASADIDIGTLICNKKPLIOESDTIVSSKSALHSSVPKSTNDROATPLSRAM--- 1260

Db 993 TDLAKGEYTNKAREN---LEEYNETDLAKGEYTNKA-RENLEEYNETDLAKGEYTN 1047

QY 1261 -----DPEKLGCDSESNSTLNSDPTVSIQDSE-----EDMIVQNSNIS 1303

Db 1048 KARENLEEYEEK---DYMNNELONGSD-GLKENALNKKELRNKSGLEKYN--ELKN 1102

QY 1304 EGFRTREOD--VEVLEPLCELVGSESTG---NCE---DRLPVKTEAGKPKPSQOKLE 1355

Db 1103 KELRNKSGDLKENLKKELONKGSBGLKENALNKKELONKQSE--GLKENALNKK 1160

QY 1356 EPPVAKCSQOIK-----LKNITDKKNE-----NRES 1382

Db 1161 ELR-NKGSBGLKENALNKKELONKGSBGLKENALNKKELONKGSBGLKENALNKKEL 1219

QY 1383 EKKGQ-----RSTQINGKDN-KPKIYLKGECLKEISESVVSGNVSPKNNINK 1432

Db 1220 QNKGSEGLKENALNKKELRNKGSBGLKENVYTNNDLKNNDIQNDLSKKDKKKNELLNK 1279

QY 1433 IIPENDIKS-----LVKESAIRPFINGDVIMEDFNERNSSEFKSHLSSSDAEGNVR 1485

Db 1280 DISNMDKKNKELLNLDLSMEDKKNELLKDIRNNDLSIGMBOQNGTKPKSGOON 1339

QY 1486 DSELEPLSTKESDSTQTTTPSASCPESNSVNOVEDMEIETSVKAVTSSPTSEESNLS 1545

Db 1340 TOLKNTPMEROQNTGLKNTPS---EGQOQNTGLKNTPSGQOQNTGLKNTPMEROQNTGLK 1395

QY 1546 N---DPIBENGEPINKNENVENESKRTVITEVTIMTSVVAETSKTYLKEGDKQTYVS 1602

Db 1396 NTPSEGOQNTGL---KNTPIBGOQNTGLKNTPS---TELKNTPS-----EGQOQNTGLK 1434

QY 1603 STEN-----CAKSTVTTTTTTTTVTLSTPSGGSVDIISVQSKVTVTVTDSITLTTG 1656

Db 1435 NTPNERQOQNTGLKANNAQOQNTGLKNTPSKG-----QOQNTGLKNTPMEROQNTG 1484

QY 1657 GLTVTSMTVSKEYSTRDVYKLMKESRPKKTSGTAL---PSYRKFFVTKSTKKSIFVLPND 1713

Db 1485 ---LKNTPNERQOQNTGLK-----NTPSEGOQNDLKNTPMEROQ-----QNT 1522

QY 1714 DLKTLARKG-----GIREVPEYFNNAKPAALDIPYSPRPPTGITWRYRLQTVKSLAGVSL 1769

Db 1523 GLKNTASKGOQNTGLKNAP----- 1541

QY 1770 MRLIMASLRMDMAKVPBGSGSTRT-----ETSETETTTETIRRDVGPYGINREYCI 1825
Db 1542 -----NEROQNTGLKNTPSEGOQNTGLKNTPSEGOQNT----- 1574
QY 1826 RKIIPIGVPEPKETPPORKGLRSSA-----LRKRPETPKQGPVILIEIHWAE 1877
Db 1575 -----GLKNTPNE--ROQNTGLKNTASKGOQNTGLKNAPNEROQNTG--LKNTPSEG 1622
QY 1878 ELEIWEIRAFARVEKE--RAQAVEQOAKKRL-----EOKPVIATSTSPSTSTST 1929
Db 1623 Q-QNTGLKNSASKGOQNTGLKNTPSEGOQNDLKNAPNEROQNTGLKNTPSEGOQNTGLK 1681
QY 1930 ISPAQ-----KVMAPISGSVTTGKMWLTTKVGSFATVTFQONKNEHOTFATWVKQGS 1984
Db 1682 NTPSEGOQNTGLKNTPSEGOQNTGLKNTPNER-----QOQNTGLKNT--PSEGOQ 1728
QY 1985 NSGVYQV---QOKVLGIIIPSSGTSGOQTFTSQPKATVTVIRPNTSSGSGTSSNSQVITG 2041
Db 1729 NTGLKNTPNEROQNTGLKNANKGOQNTGLKNTPNEG---QOQNT-GLKNTPSEGOQNTG 1783
QY 2042 POIRPGMTVIRTPLOOSTLGKAIIRTPVMVOPGAPQ---VMTQIRGQPVSTAVSAPNT 2098
Db 1784 LKNTPSEGOQNTGL- KNTPSEGOQNTGLKNTPNEROQNTGLKNANKGOQNTGLKNTPNE 1842
QY 2099 -----VSTP--GOKSLTSATSTSNIOSSASQPPRPOQGYKLTMAQLTQLTQGHGNO 2150
Db 1843 GOQNTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNANKQ---QNT 1899
QY 2151 GLTVVIOGQQTGQLQILPGCVTVLPBGQOLMAAMPNGTVQRFPLPLATITATFAST 2210
Db 1900 GLKNTPSSGOQNT-----GLKNTPNEGQ---NTGLKNTPSEGOQ 1936
QY 2211 TTTVSTTAAGTEGROSKLSPOMOVHOD---KTLPPAQSSTVG---PAKAOPOTA-- 2260
Db 1937 NTGLKNTPSEG---QOQNTGLKNTPNEROQNTGLKNTPSEGOQNTGLKNTPSEGQNTGLK 1993
QY 2261 -QPSARPOQTOPOSPAQEVOT---QPEVOTQTTVSSHVPEAOPTHAOSSKPOVAQ 2315
Db 1994 NTPNEGQOQNTGLKNTPSEGOQNTGLKNANKGOQNTGLKNTNEGQOQNTGLKNTPSEGOQ 2053
QY 2316 SQPOSNV--QOQSPYRVQ--SPQOTRIRPSTPSQLSPGOQSOVQTTTSQPIPIQPHTSLO- 2371
Db 2054 NTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNANKGOQNTGLKNT--PNEROQNTGLKN 2111
QY 2372 IPSOGOPQSPQVQSSQOTLSGOTLNOVSFSPRPOLQIQOPQOYIAPVPLQOQOVY 2431
Db 2112 TPNEGQ-----QNTGLKNTPSEGQO--NTGLKNTPSE-----GOQNTGLKNTPSEGQOQNTG 2160
QY 2432 LSQIQSQVVAQIQAOQSGVPOQIKQLPFIQIQSSAVQ---THQIQNTVTVQAASVOEOL 2488
Db 2161 LKNTPSE-----QOPNTGLKNTPNEGQOQNTGLKNTPSEGOQNTGLKNAN----- 2205
QY 2489 QRVQOLRQOQKQKQOQIEINVNTPSKLLIKVEITIQOQVMAKHNAVIEHLKOKSKMTPAE 2548
Db 2206 -----KGOQNTGLK-NTPSE-----GOQNTGLKN-----TPSE 2232
QY 2549 REENORMIWCNOVMXYIIDIKDEKQOAKKKRRESEVQKRSK--QNAFKLSALLFKHKE 2607
Db 2233 GOQNTGL-----KNANKGOQNTGLKNTPSEGOQNTGL----- 2265
QY 2608 QLRAEILKKRALDLQIEVOEELKRDLIKKEKDLQLAQATAVAAPCPVTPVLPAP 2667
Db 2266 -----KNTPSEGOQ--NTGLKNANKGOQNT-----TGLKNTP 2295
QY 2668 PAPPPSP-----PPPPGVQHTGLSTPTLPVASQKRRREKDSKSKKKMISTTSKET 2723
Db 2296 SEGQPNNTGLKNTPNEGQOQNTGLKNTP-----SEGQOQNTGLKNTPNEGQOQNTGLKNTPSEG 2350
QY 2724 KKDTKL 2729
Db 2351 QOQNTGL 2356

Search completed: November 20, 2002, 16:31:35
Job time : 142.041 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:33:37 ; Search time 13.288 Seconds

(without alignments)
3426.253 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAETDSE.....KLKGFKASRSHNNKIQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	9.3	328	US-09-925-297-816	Sequence 816, Appl
2	982.5	6.6	238	US-09-729-835-80	Sequence 80, Appl
3	443	3.0	2344	US-09-815-242-12713	Sequence 12713, A
4	438	2.9	3256	US-09-919-172-98	Sequence 98, Appl
5	432	2.9	2665	US-09-864-761-34248	Sequence 34248, A
6	402	2.7	4019	US-09-738-973-425	Sequence 425, App
7	401.5	2.7	1367	US-09-801-368-108	Sequence 108, App
8	391	2.6	5179	US-09-922-217-1068	Sequence 1068, Ap
9	391	2.6	5179	US-09-833-263-1068	Sequence 1068, Ap
10	389.5	2.6	2478	US-09-815-242-5816	Sequence 5816, A
11	389.5	2.6	2478	US-09-815-242-12967	Sequence 12967, A
12	361.5	2.4	2843	US-08-681-219-32	Sequence 32, Appl
13	354.5	2.4	1400	US-09-764-176-7	Sequence 7, Appl
14	342	2.3	2368	US-09-815-242-5635	Sequence 5635, Appl
15	342	2.3	2368	US-09-815-242-12389	Sequence 12389, A
16	339	2.3	6281	US-09-815-242-12996	Sequence 12996, A
17	329.5	2.2	5795	US-09-815-242-12610	Sequence 12610, A
18	324	2.2	1781	US-09-738-877-3	Sequence 3, Appl
19	318	2.1	1596	US-09-902-432-4	Sequence 4, Appl

20	317	2.1	1363	US-10-124-557-52	Sequence 52, Appl
21	315	2.1	2063	US-09-735-367B-2	Sequence 2, Appl
22	312.5	2.1	1140	US-10-124-557-104	Sequence 104, App
23	312.5	2.1	1404	US-10-124-557-2	Sequence 2, Appl
24	312.5	2.1	1404	US-10-124-557-62	Sequence 62, Appl
25	312	2.1	2005	US-09-735-367B-3	Sequence 3, Appl
26	311.5	2.1	1049	US-10-124-557-58	Sequence 58, Appl
27	311.5	2.1	1313	US-10-124-557-142	Sequence 142, App
28	311.5	2.1	1354	US-10-124-557-48	Sequence 48, Appl
29	309	2.1	2441	US-10-109-886-8	Sequence 8, Appl
30	306	2.0	957	US-09-922-217-1065	Sequence 1065, Ap
31	306	2.0	957	US-09-833-263-1065	Sequence 1065, Ap
32	305	2.0	1320	US-10-124-557-46	Sequence 46, Appl
33	305	2.0	1320	US-10-124-557-60	Sequence 60, Appl
34	305	2.0	1361	US-10-124-557-40	Sequence 40, Appl
35	304	2.0	1314	US-10-124-557-50	Sequence 50, Appl
36	301.5	2.0	3158	US-09-815-242-12611	Sequence 12611, A
37	301	2.0	1038	US-10-124-557-74	Sequence 74, Appl
38	301	2.0	1370	US-10-124-557-44	Sequence 44, Appl
39	301	2.0	1311	US-10-124-557-42	Sequence 42, Appl
40	300	2.0	1022	US-10-124-557-84	Sequence 84, Appl
41	296	2.0	1537	US-09-801-368-104	Sequence 104, App
42	293.5	2.0	941	US-10-124-557-14	Sequence 14, Appl
43	293	2.0	2442	US-10-109-886-10	Sequence 10, Appl
44	292	2.0	2785	US-09-801-574-8	Sequence 8, Appl
45	290	1.9	2086	US-09-815-242-5639	Sequence 5639, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-816
Sequence 816, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (178)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (269)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Db 238 SRLAMSTFASAA-----TTTALTANTTVAKNDIKOYMTTSGAATYDOSTGYVT 286
Qy 621 LESPACAGKA-SGSTRIITRLRNPDSKLSQOQVAAAHBAKFLFEKQ---EVLVY 676
Db 287 LPODTSQKALITLGTRI-----DSN---KSFHFGKVLNGLNKEHGNGNGDGIQRA 335
Qy 677 NNOGELSRSLSTKEVIMKNNINNYFKLGOEGKYRYVHNOYSTNSFLNK-----725
Db 336 FSPGVLGELTNGAAVIGLSNAEGF---KIDTYHNTSTPNSSAKAKADPSNVAGGA 391
Qy 726 -----HQHREDHDKRHLAKFCLPAG-----EKKMGVHGSVLIT 764
Db 392 FCAFVTTDYGVAAYTTSSSTADN---AAKLNOPTNNTFODEDINTNGD---TKMVT 444
Qy 765 SYLRILITOLENNIPSSFPHPWASHRAMWIKAVOMCSKPREFALALA-----812
Db 445 -----KYAGQWTNRNISDMIAK---SGTJNLSLMTASTGATNLQOY 484
Qy 813 -----ILECAVKPVYMLPIW-----REPLGH-----TRLHRTSTIE 843
Db 485 QGTEFEYTESAVYOVYVYVTTGKDIIPKITYSGNVDOYVTTIDMOQSALTAKGYNTSYD 544
Qy 844 REEKEVKKEKKEKOEETMOQATWVYKTFEPVKH---OYWKQGEYR-----888
Db 545 SGTASTYNDTKTKVKTANAGOSYTT--YFTDVKAPTYVGNQTIYEKTMNPIVLTTDN 602
Qy 889 -----VTGY-GGWSMISTKTHVRFVEKLPQNTNV-----NYRKSLEGTKNMDE 931
Db 603 GGTGYVNTVYGLPSGLSYDSATNSIIGTPTKIGOSTVTVVSTDOANNKSTTTFTTIVVOT 662
Qy 932 NM-----DESDKRKCSPPKKIKIEPDESEK---DEVKGSDAKADONEMDISKITEK 981
Db 663 TAPVYTPIDKSEVFPSPISPINIAODNSGNAVTVTGLPSGLTFEDSTNNITISGTPIN 722
Qy 982 KNOOYVELDPS--DKPCKEPEMEYDDMKTESHVNCQESSOVADVANYSEGFLHTSYK 1039
Db 723 ICTSTITTYVSTASGKNTTTFKRYEYTRNSMDSVSTSGSTQOSQVSTSKA-----D 775
Qy 1040 KTKKSKLDGLLEIRIKOFTLEEKORLEKIGKIGIKGTSTNSKNLSSESPIYTKAK 1099
Db 776 SOSASTSTSG---SMTSTASSTKSTSVLSDSVSASKSLSTSESNVSSTSTSLVN 831
Qy 1100 EECQOSMNOEOSPNA-----NNDOPEDLIQCSOSDSVLRMS-----DPSHTNKLX 1148
Db 832 SOSVSSSMGVSYSKSTSLSDFLSNSSTSEKSESVSTSDSLTSTSLSDVSMSSTSGSL 891
Qy 1149 PKDRLVDVYSIRPETKCPKONSITENDI--EEKVSDLASRGOEPTSKTKGNDFITDSC 1206
Db 892 SKQSLSSTSDSASTSOSVSDSTNSISTSESLSGSGTSEISISNSISNSVSASTSK 951
Qy 1207 LASADDIGTLICKNKPLIOEE-----SPTIVSSSKSALHSSVPKSTNDRDAP--L 1256
Db 952 LBSQSTISLSTSDKSMSTSESLSSTSDSVSGSLVAGOSQVSTSDSMSTSEMI 1011
Qy 1257 SRAMDEGLGCD-----SESNSTLENSSDTVSIODSSSEDMIVONSNESIS 1303
Db 1012 SOSMSTSGSLASDSKSMSSVSSMSTSGSTSESLSDSISTSDSKSLSTISQSGST 1071
Qy 1304 EDPFRREDVLEPLKCLVAGESTGNCEDRLPVKGTPLANKKPKSQOKLEERPVNKS 1363
Db 1072 STSTSTSSSVRME--SGSTSGSMSTSQSDSTISIS-----TSFS 1108
Qy 1364 DOIKLNTTKDKNNRESEKKGOFTSTFOINGKOKPKIYLKGECLKETESRVSAGNV 1423
Db 1109 D-----STSDSKASASTASESISTOSYST-----STSGSISTST 1141
Qy 1424 EKVNNINKIIE-NDIKSLTVKES-----AIRPFLNG---DVIMEDFERNRS 1467
Db 1142 SLTSTNSERTSTMSDSTSLSTSESDSTSDSTSDSISEALISGSGSTISLSESNSTSD 1201
Qy 1468 SETK-----SHLLSSDAE-----GNVRSLETLPTKSDSTQTTTPPASCPESNS 1514
Db 1202 SESKSASAFLESELSSESTSESTSESLSGSTSDSTSLSDSNSESGSTSTSL-SNSTSGSAS 1260

Qy 1515 VNQVEDMEIETSEKVKVTSPTITSEESNLSDNFIDENGLPIKNENVGE-----1565
Db 1261 ISPTSGSASTSVK---SESVSTSLSTSTSLSDSTSLSTSLSDSTSGSKNSLSAM 1317
Qy 1566 -----SKRVTYTEVTMTSYATESK---YIKYKEKDK-----QVYSS 1603
Db 1318 STSDSISTRSSESLASTSLSGSTSESESGSTSSSEKSDSTMSLSMSQSTSGSTSVST 1377
Qy 1604 TENCASKTYT-----TTTTVTYKLSPTSGSVDIISVKEOS 1640
Db 1378 SEESLSDSTSLSLASMSMOQSVDSNASASOASTSTSTSTSESDOSTSTSYTSQSTQSE 1437
Qy 1641 KTVVTTVTVDL-----TTTGGTLVTSMTVSKEXSTRDKVKLKFSPKTRSGALPSY 1695
Db 1438 STSTSTSLSDSTSIKSTSGOSTSTASLSGSESDSOISSTSEKSESTSTSLSD 1497
Qy 1696 RKVTYKSTKSLFVLEPNDLKLARKGIGREVEYFNYNNAKPAIDIPYSPRPTBITMR 1755
Db 1498 STSTNSGASSTSTL-----LSNBSASASESDSSSTSLSDSTASAMQSESDS-----1544
Qy 1756 YRLQYVKSILAGVSLMLRLMASLKMDDMAKVPDGGGSTRTERSETEITTEILKRRDVG 1815
Db 1545 --OSTSTSLNSOOSTSTIRMTIASESVSESTSEGSTSESTSESDSTSLSDQST 1602
Qy 1816 PYGIREPYCIRKILICIPYET-----KPEPTPORKGLRSSALPKRPREPKOTGPVY 1869
Db 1603 -----RSTASGASASTSTSDSRSTASSTSMRTSLDQSMLSLSTSTSV 1651
Qy 1870 IETWAEELMEIETAFERVEKEKAQAVEQAKRLEQAKTYVATSTSTSTSTST 1929
Db 1652 SDS-----TSLSDSVSDSTSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1689
Qy 1930 ISPAOKVNAVAPISG-----SVTTGKMYLTTTKVGS 1959
Db 1690 STASAEVMSASISDSMSSESVNDSSESVSESNSESKSMGSTSYSDSGSLSTSLRK 1749
Qy 1960 PATVTFQOKNFEHQTFATVWYKQGSNGVYVOQKVLGIIPSTSTGTSQOTFTSFOPRTAT 2019
Db 1750 SESVSEISLISGQSMDSVSTSDSSLSLAVSTQORSESESESDSLSDSKSTGSTSTST 1809
Qy 2020 V-TIRPTSGSG-----TNSQSVITGPQIRPMGVITPQOQST 2059
Db 1810 SGLSLSTSLSGSESVSESSSLSDSISMSDSTSTSDSLSGSISLSDSTSLSDSLSD 1869
Qy 2060 LGRALIRTPVMVOPGAPQOQVMTQIIRGQPVSTAVSAPNTVS-----STPGQKSLTS 2110
Db 1870 -SKSLSSQSM---SGSESTSTSVSDOSSSTISNSQFDSKISASDSKMSSTSDSSISG 1925
Qy 2111 ATSTNSIOSSASQPPRPOOQOVYKLTMAOLITQLQHGNGGLVTVIOGOQOTTGOLQIIP 2170
Db 1926 SNSTSTSLSTSDS---MGSVSVSTNS--TSLSDSTLSGS--LVSODSSSTSTSESLS 1974
Qy 2171 QGVTVLPBGQOQLMQAMPNGVQRFLETPPLATATTAATTTTSTVSTAAGTEBOROKU 2230
Db 1975 -----DSMAQOSTSTASGSLSTISLSMSASAGTLTISOSTSVSTST-----L 2017
Qy 2231 SPQMOYHOKITLPPAOSSVGPAKAPQTAQPSARQPOQOTQPSAPQPEVQOTPEVQOTQ 2290
Db 2018 STSDSISDSTISISISGQAVESSESTSDSTISIDSESLSTSGSTSTSTSTSESLSTSN 2077
Qy 2291 TVSSHVPEOPTHAAQSSKQVAAQOSQPOSNOGOGSPVFRVQSPQRIRPSTQOLSPOQ 2350
Db 2078 SGTSTV-SESLSTSGSGS-----TSVSDS-----STSSSLSTSG 2111
Qy 2351 OSQOVOTTSQPIPIQHTSLQIPSOQPOPOPOVOSTOTLSSGOTLQNOVSVS 2403
Db 2112 STSVSDSTSMSESNASISMSQISGSTSGSTSTSTSESLSMGSGTHNNTSVS 2164


```

; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Farris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO: 98
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

```

```

Query Match      2.9%; Score 438; DB 10; Length 3256;
Best Local Similarity 18.3%; Pred. No. 2e-11;
Matches 560; Conservative 418; Mismatches 1068; Indels 1010; Gaps 143;

```

```

QY 300 VPGVTDCAVEIQNKPYIRHEPIGYDSRRKRYWFLNRRLLIEDTENENKKIMY----- 354
DB 221 VP--TTCCLDNKSKNE-----SPFKLYESVKKELDYKSOKEVNIQYCKKSG 265
QY 355 -----YKRVQVLAELIDCLDKDYWEALCKLLEEMREIHHMIDTDLTKAKGSKNFL 410
DB 266 LQTDYATKEKSDAGLO-----GETOL-----VSRRKSRKSGSGGHAVA 304
QY 411 AAAN-EELESIRAKKGDIDNVKSPET-----EKDNET-----ENDSKDAENR 455
DB 305 EFAESPEDLDONKKGKRDVSVOTPSKAVGASPLYPRAKMKTPVOYSQOQNSPQKHKK 364
QY 456 EEF-----EDQSLKED---SDKTPDDDEQKSGSEPTVEGDKGNSVS--ANLGDNTN 504
DB 365 DLYTTGARESVNLKSGSGFAGDKTLPRKLTNRNTPAKVEDAADATKPEVNLSSKTRG 424
QY 505 A--TSETSTSE---GRSPVGCSEF-----PDSSNAEK-----KVASELPDQVPE 546
DB 425 SIPTDVEVLTETEHNEPEPLTLMTOVERKIQKDSLKEPEKLGTTAGQMGSLPLSSV 484
QY 547 EPNKTESSNTS-----ATTTSIQPNL--ENSSSSELNSSOSSESAKA----- 588
DB 485 DINNPODSINESBICPLAKRRVSPFGHLRPELDENLPRPTPLKRGAPTKRKSIVMHTP 544
QY 589 -----DDPE--NGERES---HTPVSTIOEELIVG-----DFTSEKSTGESESPG 626
DB 545 PVLKTIKEQPOPSGKOESSEIHVEVKAQSLVISPAPSPRKTVPASDORRRRSCKTAPA 604
QY 627 AGKGA-----GSTRIITLRMPDS-----KLSQLSKQVAAAHAHEAKLKEGEVL 674
DB 605 SSKKSQTEVPKRGGERVATCLQKRVSTISRSQHDILQITCKRRSGASEAN-----LI 656
QY 675 VVNSQGEISRLSTKK--EVIIMG--NINNYFKLQOEGKYR--VHNOYSTN-----S 720
DB 657 VASMSADVYKLGAKQOTQYKIKHGPOKSMKKRORRATPKKPYGVEHVSQSTIGHANSPT 716
QY 721 FALNKQIHRREDHDKRR--HLAKHFCLTPAGEFK-----WNGSVHSGKVLITLRL 769
DB 717 IITGKAHTEKVVHAPRPVPLNNEFISNOQKDFKEDLSGIAEMFKTPVKEQPOLT-STCHI 775
QY 770 TITQLNNITPSSP-----FHPMASHRAWIKAWQCSK--PREFALALAILLEAVK 819
DB 776 AINSNENLKGKQFGTDSGEELLPTPSESFGVGFSAQNAKOPSD-----KCSAS 827
QY 820 PVMMLPIWREFLGHNT-----RLHMTSTIERE-----EKEK 849
DB 828 P-----PLRRCICINENGNAKTPRNTYKMTSLFETKTDTEPESKYVSTVNRSGRSTEFNR 883

```

```

QY 850 VKR--KEKKOESEETM-----QCATWVKKYTFPVKHQVWVKQKEGEVRYTGQNSWI 898
DB 884 IQKLIPESKSEENTLEVEICILKRGOKATLLQ-----KREGMKEL----- 925
QY 899 SKTHVYRFVFKLPQNTNVNRYKSLGKTKNMDEMDESDK-----RKCS--RSPKTKIE 951
DB 926 -----ERPFEET---YKENIELKEN--DEKMKAMKRSRTWQKCAPMSDLDLTKSL 970
QY 952 PDSE--KDEVKG-----SNAAGADQENMDIKTEFKKQDVKELLDSQDPCKE--EP 1002
DB 971 PDTLEMKOTARQONLQOTDHAAPKSEKGIYK-----PCOSLOP 1012
QY 1003 MEYDDMKTESHVNCOESSQOVYVNVSE-----GFHLRTS-----Y 1038
DB 1013 -----EPINTPTHTKQQLKASLKGVGVEELLAVGKFLRTSGETHHREPADGQKSLRFP 1068
QY 1039 KKKTK-----SKLUDLLEPRIKQFTLEEKORLEKIKLEGIGIGTSTNSSKNLSESP 1093
DB 1069 KESPKQILDEPAARVGTG--KKWPRTPKEEAQSLEDL---AGFKELQTPQPSSESMTEK 1123
QY 1094 VITKAKGCSQDSMRQPSNANNDOPE-DLIGQCSQSDSVLRMSDPSTHTNKLVPKDR 1152
DB 1124 T---TKIACKSPPEESVDITSTKQWPKRSLKRADEVEEFLAIRKLTPSAGKAMLPKPA 1180
QY 1153 VLDDVSI-----RSPETKCPKONSIN-----DIEEKVSDL 1183
DB 1181 GDEKDIKAFMGTPVQKLDLAGTLPGSKROLQPKERQAQLELAGKELFQTPGHTEEL 1240
QY 1184 ASRGQ-----EPTKSKTKGNDPFIQDSKLASADIGTLI--CKMKKPLIQE 1227
DB 1241 VAAQKTKTKIPCDSPQSDPVDPTSTKQR-----PKRSIKAVEGSLACRNIMP--- 1290
QY 1228 ESDPTIVSSKSAHLHSVPKSTNDNR-----ATPLSRAMDEGKLGCDSESNSTLNSDPT 1282
DB 1291 -----SACKAMHTPRPSGEEKDIIIFVGTGVQK-----DLTENLUGS 1329
QY 1283 VSIQDSSEDMIVQNSNESISEQFRTREODVEVLEPIKCELYSGESTG--NCDRLPVKGT 1341
DB 1330 KRRPQTPKEEAQALEDTLGEKELFQTPGHTEEA-----VAAQKTKTKMPCESSPESAD 1382
QY 1342 EANGKKPSQOKLEERPVNKCSDQIKLNTDKNNENRESEKKGQRTSPFOINGDNKP 1401
DB 1383 TPTSTRQOPTLEKROVQELSALK-----KLTQSGETHHDKVPGGEDS 1430
QY 1402 KIYKGECLKEISRYVSNV--EPKVNINKIIPENDISLTVKGSARLFPINGVIME 1460
DB 1431 INAFREAKOKLDPAASVYSGSKRHPKTK--EKQPLEDLAGW--KELFQPR-----VCI 1480
QY 1461 DFNERNSETKSHLSSDAEGVYRDSLETLPSTKESDSTQTTTPSASCPESNVQVED 1520
DB 1481 DKPTTHEKTKTKIACRSQPD-----PVDTPSSKQSKRSRLKYD 1519
QY 1521 MEIETSEVKKVTS-----PTSEESNLSPNDFIDENGLPIK--NEWNVNGESKR 1568
DB 1520 VEEEFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPVQKLDLTELUTGSKR 1574
QY 1569 KTVITEVTTMTSTVA-----TESKTVIKVEGDKQTVVSTENCASVTYTTT 1616
DB 1575 LQTPKEAQAQLELADLAKRELFOTRGHTESMT---NDKTAKVACKSQPOLDKNPASSKR 1631
QY 1617 TTVTKLSTPSTGSSVDIIISKEQSKTYVTTVDLSLTGGLVYTSKYESTYDQYK 1676
DB 1632 RLKTSLG--KVGKVEELLAVGLTQSGETHHTHEPTGCGKSMKAFMESPKOILDSAS 1689
QY 1677 LMKFSRPKTRGSTA-----LPSYRK--FVTKSKTKSLFVLPNDL----- 1715
DB 1690 LUGSKROLKTPPKKSEVPEDLAGFIELFQTPSHTKSMNTEKTKVSYRASQPDLDVTP 1749
QY 1716 -----KTLARKGIGIREVPYFNANAPALDINWYPSBPFPGITWYRLQTVYSLAGVSL 1769
DB 1750 SSKPQKRSRLKADTE-----EFLAFRQOTPSAG-----KAMHTPKPAVGEK 1793
QY 1770 MLR-LIMASLRMDMAKAVPGGGSTRTETSETETITTEIIRRDVGPYGIR----- 1820

```

Db 1794 DINTFGTGVOKIDDPGNLP--GSMRRLOTRKKAQALEELT-----GFRLEFOTPCT 1844
QY 1821 -----FEYKIRLIC--PLGVP-ETPKERTPOKGLRSS-----ALRPKRP----- 1859
Db 1845 DNPETDEKTKTKLCKSPQSDPADPTNTKORPKSLKADVEEFLAEFKLTLPAGKAM 1904
QY 1860 EMPKOTGPVLIETWAVEELEMEIRAF-----AEVE-----KEKAQAVE 1900
Db 1905 HHPKAA-----VGEK-----DINTFGTGVPEKIDLGNLPKSRRPOTPKKAKALE 1952
QY 1901 QAKKRLLEOKPLVIATSTTSPTSTTSPAQKVMVAPISGVTGTMAVLTTVGSP 1960
Db 1953 DLAGEFELQOTPOHTEESMTDKITEVSCSPQDPVKTTP-----TSSKORLIKISIGK- 2005
QY 1961 AAVTFQONKNFHOTFATFWKQOGSNGVQYQOKVL--GIISSTGTSGOTFSPQRTA 2018
Db 2006 -----GVKKEVLPVGLTQTSKTTQT----- 2028
QY 2019 TVTIRPNTSGSGGT-----TSNSQVI-----TGPOIRPQMTVIRTP-----LOOSTLG 2061
Db 2029 -----HNETAGDGKSIKAFKESAKOMLDPANYGTGMERMP-----KTPKEAQSLDELACF 2079
QY 2062 KAIIRTPVMVQEPAPQOVMTQIIRGQPVSTAVAPNTVSSTPGOKSLTSTSTNIOSSA 2121
Db 2080 KELFOQTPDHTHEESTTDDKTKTKI-----ACKSPPESDMTPTST----- 2117
QY 2122 SOPPRQOGQVOKLT--MAOLITOLTOGHGNGQLTVVIQGGGQOTGLOLPL-----OGVTV 2175
Db 2118 RRRPKTPLGKRDIVEELISALKQLO-----THTDKVGGEDGGINV 2159
QY 2176 LBPQGOOLM-----QAAMPNGTVQ-----RFLF--TPLATVATTASTTTTV 2215
Db 2160 FRETAOKLDPASVYSGKQOPRTPKGAQPLDLAGIKLFLQTPICOTDKPTTHEKTKI 2219
QY 2216 SITTA-----AGTGE--QROSKLS--POMVHQDKTLPPAOSSSVAPAKOPQA----- 2260
Db 2220 ACSRPOPDPVGTPIFKPQSKRSLKADVEEELALRKRTPSVKAMDTPKPGAGDEKDM 2279
QY 2261 -----QPSARPOPOTQPOSAPQEPVOTQPEVOTQTVSS-----HNPSEAOPT 2303
Db 2280 KAFMGTPVOKLPLGMLPGSKRMPQTPKE-----KAQLEDLAGEKELFOTPGDKPT 2332
QY 2304 HAQSKPOVAASQPOSNNVQOGSPVEVQSPQTRIRPS-----TPS- 2344
Db 2333 -TDEKTKTACKS-PQ-----PDPVDPASIKQKRPKRLKADVEEELALRKRTPSA 2383
QY 2345 -----QLSPGQOSQVOTTSQPIPIOPHTSL-QIP-SQGOPOSQOP----- 2382
Db 2384 GKAMDTPPKPAVSDKKNINTEFETPVOKLIDLGNLPKSRQPOTPKKAEALDELVGFKEL 2443
QY 2383 -QVQSTOTLSSQOTLNOVSVPSSRPAQIQOPQOVIAVP-----QLOQOVVLQSI- 2435
Db 2444 FQTPGHTESMTDKITEVSCSPQESFTSSKQRLTPIPLVKVMEKPLAVSLTR 2503
QY 2436 QSQVVAQIOAOSQVPOQIKL--QLPTIOQSSAVOTHOIQNVTVQAASVQEQOLRVOQ 2493
Db 2504 TSETQTQTHTEPLGDSKSIKAFKESPKQILDPAA-----SVTSRRLRTRKKEKARA 2555
QY 2494 LRQOQOKK-----QOQIETNVN-----TPSKLILKVEITIQOVVM 2529
Db 2556 LLEDLVDFKELFSAFGHTESMTDKNTKIPCKSPPELDTATSTKRCPTKRL--KREVK 2613
QY 2530 KHAVAIEHLKOKKSMT-----PAEREENORMIVCNOVMYIILDKIDKEKQAKKRRRE 2583
Db 2614 EELSAVERLQTSQSTHTHKEPASGDEGKIVL-----KQAKKPPNP 2656
QY 2584 ESVEOKRSKONA-----TKLSALFKHKEQLRA-----ELKK 2616
Db 2657 VEPEPRRRPRAPKAKAQLPLEDLAGEFTLESETSGHQSGLTACKATKIPCESPLEVDT 2716
QY 2617 RALDDLOLQIEVQEEIKRDLIKKEKDQLOAO-----AT 2651

Db 2717 TASTKRLHRTVOK-----VQVKEPSAVKFTQTSGETTDADKEPAGEDKIGALKESAK 2771
QY 2652 AVAAPCPVTPVLPAPPAPEPSP-----PPGVQHTGLSTPTLPAVASQKRRREE 2703
Db 2772 QTPAPASVYSGKRRRRAPRAPSQAIEDLAGFKDPAAGHTESMTDKTKTKICKSPSEL 2831
QY 2704 KDSKSKSKK-----KMISTSKETKDTLYCICKTPYDESK 2741
Db 2832 EDATSSKRRPRTRAQKVEYKEELLAVGKLTQTSGETTHND-----KEPVGEK 2880

RESULT 5
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14

```

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; OTHER INFORMATION: EST_HUMAN HIT: A0117052.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALU0 3.00e-10
US-09-864-761-34248

```

```

Query Match      2.9%; Score 432; DB 10; Length 2665;
Best Local Similarity 19.1%; Pred: No.2.9e-11;
Matches 547; Conservative 397; Mismatches 1112; Indels 810; Gaps 131;

```

```

QY 7 EEDGDAEETQDSEDD-----EDMEEDDDSDYPEMEDDDDDASYCTESSFRSHS 59
D 170 KNEITDKERTFDEPERVERERRLIRKEKVEKDKTKQ----- 205
QY 60 TYSTPRRRPRVHRPRSPLE-----EKDIPLEFPKS-----SEDLMPNEHI 104
D 206 -----KRGKGVHSPSSQSSSETDQENERSQSP-EKPRSCNKLREKADKEGIAKNLEL 257
QY 105 MNVAIAYEVNRNFTVRLSPFRFEFCALVSOQCTTAAEMHVLLKAVLREEDTSNT 164
D 258 MCVVLTFRVKEKGVLDHTRP-----EKIKALMDTYKSSALDOKLOYST 305
QY 165 TRGPADIK--DSVNSTLYFDGM-TWPEVLRYCESD--KEYHVLPRYOEAEDYRGPV 218
D 306 EPARKSDLSKLESYRMKPKKESLSHSEVY-----EKEGRLAKRHLKPEQPAD-----GVS 357
QY 219 ENKIVLQFLVDQFLTNIRBELMSF---GVIOYDHCHVCHKLDDLCCEFCSAVYHL 275
D 358 AVULEKLEAKRRPRADSNLAEKQKPEVKSSPEMED-AVLSKSKOPDVSSREVILLREG 416
QY 276 ECVKPRPL--BEVDEMOCEVCVAHVPGVTDCAVETOKKRPYIRHEPIGYDRSRKRYM 333
D 417 EABRKIPVRKRIILRESKIKILDLNTVYASPKDC-QELAS-----ISVGSGRSSD 466
QY 334 LNRRL--IIEPTENENEKTIWYSTRKVOALDIDLCDMDYEAELCKIIEEN--REET 388
D 467 LQARLELGAESVENOEVOSKRPDKPRKPOLKQ-----QVLDGDPGPREVY 512
QY 389 HR-HMDITEDLTNRKAGSNKSFLLAANEETLESIRAKG-DIDNVKS--PEETEDKNE 443
D 513 RKVYCLARDETPRKSGQESHVNTBEKI-----GIDIDHTOSYRKQMOQSRKQO 564
QY 444 TENDSDAE-----KNREEFEDQSL-----EKSDDKTPDDPEQKSEPTFVG----- 488
D 565 MEMEIKASEKFGSPKRDVDEYERSLVHEVGRPOVDVDSPPSKKRRMDHVDICTKR 624
QY 489 DKGNVSANLGDNTNATSEET--SPS-----EGRSVYGC--LSETPDSSMAEKVYA 537
D 625 ERNRYRSROJSED-----SERTGSPSVRHSFHEDEDPISPRLLSVKSPGVYDK-- 676
QY 538 SELPQD--VPEEPNK--TCESSTNTSATTTSIQPNLNSNSSELSNOSSESAAK-- 588
D 677 -VLPYNTIVREBSLKFNFYDSSRRQMDMAKIKILSVLNSBELNKRMSQAKQDAGR 735
QY 589 -----DPPENGERESHPTPVSIQEIIVGFTSEKSTGE 620
D 736 VSEPNIIIRDSLARKRSVLDLEPGEVPSDSEDEGKHSHP-----RASA 780
QY 621 LSESPAGAGAGASTRIILRLNRPDSKLSQCVAAAHAAHANKLFEKG-KEVLVNSO 679
D 781 LIE-----SSRLSTLLDRDREDKLRE-RDERLSSSL-ERNKFYSFALDKTITPDNK 828
QY 680 GEISR---LSTKKEVIMKGNINNYFKLQEGKRYVYHNOY----- 716
D 829 ALLERKAKSLSSSR-----EMKSLMDSDSRFANFRNKKDKKEKYDASAPRPLPSWTMKK 881
QY 717 -----STNSFALNKQHREDHDKRHL-AHKFCLTPAGEFKWNGSVHGSKVLLITSLRLT 770
D 882 KIRTDSEGMKDDKEDHKDEQEOELFASRF-----LH-SSTIEQDSKRI- 926

```

```

QY 771 ITQLENNIPSSFF-----HPNMAISHRANWIKAVOMSKPREFALALALECAVKNVMLP 825
D 927 -QHLEKREEDSDTISGLIYKQISEGAN-----STDSIOEPVLLF- 966
QY 826 IMREFLGHTRLHMTSIEREKEKVKKKEKKQOEETMOATWYKYTFPVKHQVWKQKE 885
D 967 -----HSRFMELTRMQQKEKEDQKPEVEKOEPTENHP-----KTPESAPENKOS 1012
QY 886 EYRV---TGYGWSNISKTHVYRFVPRKLPNGTWNVNRKSLSEG---TKNNDEMMDSDKR 939
D 1013 ELKTPPSVGPSTVYVTLVESAPSALEKTTGD-----KIVEADLVTEKTEPATVSEBA 1066
QY 940 KCSRSFKIKIIEPDSEKDEKVGSDAAKA-----DQ----- 970
D 1067 KPASEPAPAVPEQLDQVDLPRGADPRKEAAMMAGVEBSSGQPPRYLDAKPTPGASPS 1126
QY 971 -----NEMDISKITEKKDQVYKELLDSDSKPCKEEMPEVDDDMKTESHVNCQ---E 1019
D 1127 QAESNVDPPEPDSYQPLSKPAQKSEENEPKAEKPDATAEDPANOKAALAEPSQPAE 1186
QY 1020 SSGQDVVNVSEGFHLRTSYKKKT-----KSKLDGLLEBRRIKQFTLE-- 1061
D 1187 DLEVDPVPAKDKKPKKSKRSKTPVOAALAVSIYEKFTKRSKID--REKLKRSNPRG 1243
QY 1062 EKQRLERIKLEGGIKIGKSTNSKNLSESPVYTRAKEGCOSDSMKROQSPNANNQPE 1121
D 1244 EAQKLELMEKA--EKITTRASKNSAADLEHPEPSLPLSTRRRNRVRSYVATMGHENS 1301
QY 1122 DLIOGSQSDSSVLRKMSDSHTINKLYPKDRVLDVST-----RSPEYKCPKONSIE 1173
D 1302 PVKEPVEQ-----PRVYTRKRL--ERELOEAAVPTPRRGRPKPTRRADDEEBE 1348
QY 1174 NDIEEKVSDL--ASRGOEPTKST-----KG-NDFFIDSKLASADDIQTLICKNK 1221
D 1349 NEAKEPAETILKPEGRMSRSOKTAAAGGPOGKKGNEKVKDATREATTETGPOIGVKE 1408
QY 1222 KPLIOESDTIVSSKSAHSSVPKSTNDRDATPLSRANDFGKLGCDSESNSTLSSSD 1281
D 1409 SSMERPAAEAEAGSEQKRRDKDAGTDKNPETAPEVY--EKKPAPEKNSKSKGRSRN 1465
QY 1282 TVSIODSFEEDIMVONSNSIS-----EOPRTREDQVLEPLKCELVGSESTGCEDRL 1336
D 1466 SRLAVDKSAS--LKNVDAVSPRGAAAGAESEGVAVSPESKE----- 1508
QY 1337 PVKGTENAKKPSQOKLEERPVNKCSDQIKLKNITTDKKNNEKRESEKKGQSTFQJNG 1396
D 1509 -----SPQKE-----DGLSSQLKSDPVDPDPKPEPEKEDVSASGSPPEATQL-- 1548
QY 1397 KDNKPRITYLKGECLAKETISRVVSGVNEPKVANNKIIPENDIK-----SLTVKESAIRP 1451
D 1549 ---AKOMELQAVEHIAKLAEASASAAKADAPEDIAEDBKPAHQASLETALAAIGS 1604
QY 1452 FINGQVIMEDFN-----ERNSEKSHILLSSDAG 1482
D 1605 IIN-DISGEPEFPAPRPYPPGSGQTDLPAPGAQALQSEBEMEDAEVSGILEEATTE 1663
QY 1483 NYRDSLETU-PS-----TKESDSTQTTTPSASCPSNSVNYQVEDMEIETSEKKVYSSPI 1536
D 1664 SSRPPVAPNADPSAGPTDTREANGNSSET-SHSVPEAKSGKEVEYLVKKDKROKTT--- 1719
QY 1537 TSEBSNLSNDEIDENGFLPINKNEVNGESKRYVITYETVNTSTVATSEKTVIKVERGD 1596
D 1720 RSRRRKNNMKKVAVAVESHVSPESNOAGES--PANEGTVOHDEADQOE-----EKOS 1770
QY 1597 KOTVAVSTENKAKSTVTTTTVTYKLTSPSTNGSVDIISYVEQKSTVYTVYTSLLTTG 1656
D 1771 EKPHSTPPQSC-----TSDLSKI--PSTENSSQELISVEERTPT--KASPPDLP-- 1817
QY 1657 GTLVTSMTVSKRYSTRDKVYKLLKFSRPKRTSGTALPYSYRKFVTKSTKSLFVLPNDLX 1716
D 1818 ---PPAVVDEPPQARFVYHSTIESDPVTTPSDPSIP-----IPTLPSYIAA 1861

```

```
QY 1717 KLA---RKGIFREYFNYNNAKPAIDIMYPSPRPFGITMRYLQTYKSLAGVSLMLRL 1773
Db 1862 KLSPPVASGGI-----PHOSP-PTKYTEMITRROEPEAOSTPS----- 1898
QY 1774 LMASLRMDMAKVPBGGSFTR-----TETSETETITTEIKRDRVGYGIRFEYCIRK 1827
Db 1899 --PALPPTKASDVOTSSSTLRKILMDKRYSATVSTISV-----TTA 1940
QY 1828 IICPIGVETPEKETPTPOKRLRSALRKRREPRTKQGTGPVLIETWVAEELELEIRAF 1887
Db 1941 IAEVSAAPCLHEADPP-----PVDSKRLPEKTAAPV-----TNNSEIOASEVLVA 1987
QY 1888 AERVEKEKAQAV-----EQOAKRLLEQOKPYIAISTISPTSTST 1929
Db 1988 A---DKEKVAPIAKITSVISRMFVSIDLENSOKITLAKPAPQTLTGLVASLTLVWNS 2044
QY 1930 ISPAQKVVAVPISGVTGTTKVLTLTKYGSFATVTFQONKNHGFATWVKGOGSN--SG 1987
Db 2045 LVPV-NALKGPKGVSVT-LKSLVSTPRAG-PVNVL-----KGPVNVLTG 2085
QY 1988 WVQV---QOKVLGIPSTSTGTSQOTFSOPRTATVIRPN--TSGSGTTSNQVIT- 2040
Db 2086 PVNVLTTPVNAVTGTVNAPGVNNAASAVNATASAVTYTGAVTAASGVTATGTVM 2145
QY 2041 -GPOIRPGMTYIRTPLOOSTGKALIRPVNVYQCAPQOQVMTQIIRGQPVSTAASAPNTV 2099
Db 2146 AGAVTAPS---TKCKQ---RASANENSREHPGS---MPVIDDRPADAGSAGLHV 2191
QY 2100 SSTPQKSLTSTSTSNIOSSASQPRPQOGQVTKLMAOLTLQJHGNOCGLTVYIQO 2159
Db 2192 NTSESVLIS-----YSGQKTEGFOR---ISAKISOIPRA---SAMDIEFO-- 2231
QY 2160 GOTTGOLILIPQVTVL--PCPGQOLMOAAMPNGVORFLPLATTAATTATTTT--- 2213
Db 2232 -QSVKSOVKPDSVTASQPRPSKGPQ---APAGYANVATHSTVLVLTQVTYASVYISSV 2285
QY 2214 -----TVST--TAAGTGE-QROSKLSPOMOVHQDKTLPPA---OSSSVG 2251
Db 2286 KADRPSEKREPFIHLSVSTPYOGGTGVKVLQGINTPRVLYHNOVLPLPSIYTTNKKLAD 2345
QY 2282 PA--KAOPOTQPSARPQOPQOPQSPAP-----EVQNPREVQOTQTVSSHVSEAO 2301
Db 2346 PVTLKIEFKVLOPAPNLGSLTPHHPALPSKLTPEVNHVPSGSPSIPADRTV-SHL--AAA 2402
QY 2302 PTHAOSKRPVAAQOSQSNVQOSPVAVOSBQTRIRPSTPSQSLSPGOOSQVQTTTSP 2361
Db 2403 KLDASPPR-----SGRPSSEPRAS---HPSSASTALSTNATVMLAAGTP 2446
QY 2362 IP-----IQPHTSLOIPSGOPQ-----SQPOVOSTQTILS-GQTLNOVSVSSPSRP 2408
Db 2447 VPOFISIHPEOSVIMPHSITQVLSLSLSQGEVAMNTPLPSITYSIRPEALHSPPAP 2506
QY 2409 OLOIQO-----POPOVIANPOLQO-----VOVLSQIOSQVY---A 2441
Db 2507 -LQPOQIVVARPQRASTQOPAPAGVAPALASQHPREEVHYHLPVARADAPAVOSEVLVMS 2565
QY 2442 QIOAQOSGVPOQIKLQLEPIQIOQSSAVQTHQIONVTVVOAASVQO 2487
Db 2566 EYRLHPYVPRDVIRIMVPHVTAVSE-QPRADGVYKVPSPASKAQO 2610

RESULT 6
US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Radooh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
```

```
; APPLICANT: Inditias, Carol Joseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738.973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match      2.7%, Score 402; DB 10; Length 4019;
Best Local Similarity 17.3%; Pred. No. 8.8e-10;
Matches 510; Conservative 314; Mismatches 992; Indels 1126; Gaps 108;

QY 254 CVCCHKL---GDLCCETCSANYHLECYKRPLEVEPEDEMOCEVCAHKVPGVTCVAEI 310
Db 65 CACGKATDPGRLLLCDDCDISYHYCLDPPLQTPKGGKWCRCNGCATAGLRC 124
QY 311 QNKKPYIRHEPIGYDRSRKKYFWLNRLIEEDTENENEKKIMYVSTKVQLAELIDLDK 370
Db 125 EMONNTYQCAPCA-SLSSCPVCYRNYR---EEDL-----LIQCRC 161
QY 371 DYWEALCKLIEEMBEELIHRHMDITEDLTNKAQSGNKSFLAANEELIESIRAKKGIDN 430
Db 162 DRWMAVACQNL-NTEBEVENVADIGFDCS---MCRPYMPAS-----N 199
QY 431 VASPEETEDOKRETENDSDAKAKNEEPEEDOSLESDSDDKTPDDDPQCKSEPEPEYGDK 490
Db 200 VPSSDCE-----SSLVAQIYTKVELDPPTYQDGV----- 232
QY 491 GNSVANLGDNTTNATSEETSPSEGRSPVGCSETPPDSNMAE-KVAASELPQDVEEPN 549
Db 233 -----CLTE---SGMTQLOSLVTVPRRRSRPK 258
QY 550 KTCESN--TSATTTISIQPLNLSNS--SSELNSSQSESAKAADDPENGERESHTPVSIQEE 607
Db 259 LKLIINQNSVAVLQTPPDIOGSEHSRDEMDSDRGELMDCD---GKSES---SPERE 310
QY 608 IYGDFTSEKSTGELSESRGAGGAGSGSTRITRLNPNDSKLSQVAAAHAKNLF 667
Db 311 AYVDET-----KGVGTDGVKKRKRKPP-----Y 333
QY 668 KKGKVLVNSQGEISRLSTKKEVIMK---GNINNYFKLQGBKRYVYHNOYSTNSFALN 724
Db 334 RFGIGFMYRQBSRTGQGKTKRSVIRKDSGSIISQLPCRDG----- 376
QY 725 KHQREDHDKRHLNAKFCLPAGEFKWNGSVHGSKVLITSTRLITLTOLENNIPSSFEH 784
Db 377 ----- 376
QY 785 PNWASHRANWIKAVOMCSKREFALALAILBCAYKPVVMLPIWREFLGHTRLHRTSTER 844
Db 377 -----WSQOLDPTLVDESVSV-T 393
QY 845 EKEKVKKKKKKOE--BET-----MOQATWVYKTFPVKHQVKKQGEYRYTGVGSGWMI 898
Db 394 ESTEKIKRKRKKKKLEETPPAYIOEAFPGKDLADTJRQ-----SKI 436
QY 899 SKTHYRFPVKLPNTNVNRYKSLSEGTNNMDENDESDDRKCSHPKIRIIEPSEKDE 958
Db 437 SLDNLSDEGADQLVYTKNN-----TGPLDPSLD--PLSSSSAPRK----- 475
QY 959 VKGSDAAKADONEMDISKITEKKDQYKELIDSDSPCKEPEEPK-EVDDDMKTESHYNC 1017
Db 476 ---SGTHGPADDPADLADISEVL-NTDDDIIGIISDILAKSVSHSDIGPYTDDPSSLPQPNV 531
```

QY 1018 OESSQVYVNVSEGFHLRTSYKAKKTSKLDGLLEERRIKOTLEEKORLEKILEGGIKG 1077
 Db 532 MGS-----RPLSEQLDGL-----SPELDKMTDAILG 562
 QY 1078 -IGKTSTNSKULSESPVITKAEGCQSDSMROESP-----NANNDQEDLLIOGCSO 1129
 Db 563 KTKKIPELGGKVEDLEFTAVLSPANTQPIPLPQPPPTOLLPIHNDADFSEMLMGLIG 622
 QY 1130 SDSEVLRMSDP-----SHTNKLKPKORVLDVDSINSPEKCKON-----SIEND 1175
 Db 623 SPSHLHNSLIPGSGIGTSAIAQSSYDPAR--DKNSAFPMASDPNNSWTSSAPVEGE 660
 QY 1176 IEEKVSDLASRGCEPTKSKTGNDFIDSKLASADDIGLICKN-KKPLIOESDPTIYS 1234
 Db 681 -----NDTMSNAORSTLKMEK-----BEALGEMATVAVLYTNINFPRLKEEFPDWT 728
 QY 1235 SSKSAHSSVPKSTNDRDATPLSRAMDFEGLGCD--SESNSTLEMSDPTVSIQDSSE-- 1290
 Db 729 RVKQ-TAKLWRKASSQERAPYVOKARDNRAALINKVOMNSDMKROOQDSDIDPSSRID 787
 QY 1291 ----EDMIVONSNESISEQFRTREO-----DVEVLEPLKE----- 1322
 Db 788 SELEKROPLKORESEH-EQEKKFRQOMKQSKQOAKIETQKLEFOVKNEDQOQOQOQPGSQ 846
 QY 1323 ---LVSGESTGNCEDRLPVYKGTANGK-KPSQO-----KLEERPYNKCSQOIKLKT 1371
 Db 847 HLIVQSGSDPTSGISQSPILPQPGNGNMSPAQFHKLEFTKQPPSPITSTSDVAVK 906
 QY 1372 T-----DKNNNRESEKKQOTSTFQINGKDNK-KIYLK----- 1406
 Db 907 APPPPAPSKPIPIODLSQAQTSQPPSPQVSPGSSNSRPPSPMDPAKAVGTPRPPV 966
 QY 1407 -----GECLEKISESHVSGVNEPKVNNINKIIPENDI--KSLTVKESAIRP 1452
 Db 967 HFSRRNSAARVENCPTPLSSVRPLOM-----FTANRSPVADLCSSSTTNNDPAK 1022
 QY 1453 INGDVIMEDENERNSESTKSHLSSDAEG-----NYR 1485
 Db 1023 DTPRPVMTQDPFKSLGSRPVVSEQTAKPIAGTSDHFTKPSPADVQORQIRIDSYA 1082
 QY 1486 DSELTLPSTKESDSTQTTTSPASCPES-----NAVNOV-EMETIETSEVKVTSPTSEE 1540
 Db 1083 RPLLT-PARLDGSPGPKPIPMOPPPSSQDPYGSQASRRLSVDERPALTRPINDFS 1141
 QY 1541 ESNLSNDFIDENGLPIKNKNENVNGESKRTVITEVTMTSTVATESKTVIKVEGDKQV 1600
 Db 1142 H-NQSDNPSQ--PLPHRAVN-----ESFAHPSRAFSQCGTISRPTSQDPYSQ 1188
 QY 1601 VST-----ENCAKSTVTTTTTIVTKLSTPSTGGSVDIISVEQSKTVTTVTDSLT 1655
 Db 1189 PPGTPRPVVDYSQSSGASRNTDPSQPGCTPRPTTVDPYSQOQTPRSTQDLEFV 1248
 QY 1656 GGLVLSMVMYSKRYSTRDKYKLMKFSRPKTRSGTALPSTRKVTYSTKSIIVLNDL 1715
 Db 1249 ----VTNORHSDPIAHPPGTPRGISVPSQAPATPRPISSEGTSSMTPLVMPOD- 1303
 QY 1716 KKLARGGIREVYFN--YNAKPALDIMPPSPRPFGLTWRYRLQTVKSLAGVSLMLRL 1773
 Db 1304 -----PFLQAOONRGPAL--PGLVLR----- 1323
 QY 1774 LMASLRMDMAKVPVPGGSGTRETSETITTEIIKRDVGYGIRFEXCIRKTIICPTG 1833
 Db 1324 ----- 1323
 QY 1834 VPETPKETPPQKKLRS--SALRPKRPEPKOTGVIIETVAEELBELMETRAERK 1891
 Db 1324 -PPTCSQTPPPGGLDSTFSTRVSPSAARDPYQSPMTPRS----- 1363
 QY 1892 EKEKAQAVEQOAKKRLQOKP--TVIATSTSTSTSTSTISPAQVMVAPISGVTG 1948
 Db 1364 -QSDSGTGTOTAHVADQRPFGSGSGFCASSNSPMHSQOQFSGVSO-LPGVPTSGVD 1421

QY 1949 TKMVLTKVGSFATVFEQKNKNEHOTFAFWKQOGSGNV---QVQKVLGIIPSSGT 2005
 Db 1422 T-----QMTVMAQADTEKLRQKRLREIILQOQOQKIKAGROEGSOD 1465
 QY 2006 S-----QOTF-----SFOPT 2017
 Db 1466 SPAVHPGRLQHWOPENVNOATFRPPPPYPGNRSFVAPPLGRVAVFPKDOGPPPV 1525
 QY 2018 ATVTIRPN-----TSGSGTSSNS-----QVTFG--POIRPGMT----- 2050
 Db 1526 ASMGMRPHGRFPFGSGHGTMSQERFLVPRQOIGSGSVPLRRKSVDMDRPLANSQ 1585
 QY 2051 IRTPL-----QOSTLGKAIT-----FRVWVQGA----- 2075
 Db 1586 MNPVGLPQHFSQSLPVQOHNILGOAYIELRRARDGRQLFFSAPPSVEASSNLRH 1645
 QY 2076 -----POQVNTQI-----IRGQPVSTANVTY- 2099
 Db 1646 GNFIPRPDPGPRHNDPMRRPQGLPNQLEVHPDLEQVPPSOEOGSHVSSMVKRTLN 1705
 QY 2100 -----SSTPGQSLTSATSTSNIOSSA-----SOPRRQOGVYK----- 2133
 Db 1706 HPLGSESEARPLSTSVPSSETSDNLQITTPSGLEKLDSDPSVKELDVKLEGEVYK 1765
 QY 2134 -----LTMALTLQVGHGNOGLTVVIOGSGOTTGOLILP----- 2170
 Db 1766 DLDEDELENLNDTEDEKVEIDTLNLETNENLDLRL-----SGEDDIAVTPDELD 1820
 QY 2171 -----OGVTVLPBGQOLMQAAMPNGTVORPLETTPATTA 2208
 Db 1821 MCKKSMFNEELDLPIDDKIDNOCSVEPKKQO-----EKTLVLSKHSPOK 1870
 QY 2209 STTTTIVSTTAAGTEGROSKL-----SPOMVHODKTLPPAOSSSVGA-- 2253
 Db 1871 STVTNEKTEVLSPNKSVEKCEKENDENKNDVPCQASASHDLMGCKEKLHPCP 1930
 QY 2254 -----KAOPOTAOPARPOQTOPQSPADPEVOTQREVTQTTVSSHVSEAOPT--HA 2305
 Db 1931 DLEKRTNRETAPGSAN-VIOASTOLPAQDVINSCGITSGTVPILSLANEKSDNSDIR 1989
 QY 2306 QSKPQVAAQOSQOSNVQSGPVVQSPQTRIRPSTPSQSLSPGQ-----QSO 2353
 Db 1990 SGSPPEPTLPASSNHVSLPPF-----IAP-----PGRVLDNAMSNTVVS 2033
 QY 2354 VQTTSQPIPIOPHTSILOIPQ-----GQPOSO--POV-QSSQTQLSSGOTL--N 2398
 Db 2034 VNHVFSQVQVNVGL--IPGQSTVNHSLGCTKPRATQGTQSGTSSMSGPOQLMTP 2090
 QY 2399 QVSVSSPRQOLOIQOPQOVIAVPOLOQOVYLSQIOSV----- 2439
 Db 2091 TLAQONRERPLLEQPLLDLDLDEROEOOQORQOMAIRORSBPFPENIDFADITDP 2150
 QY 2440 -----VQIOAOS--GVPOQIKLOLPIPIOSSAVO----- 2469
 Db 2151 IMKAKWALKGINKVMAONLGMPPWMSRFPMPGOVVTGTONSECONIGPOAIPDDGSI 2210
 QY 2470 THQION-----VTVQOASVOEOLQVQOOLRDOQKKKQOQIETINVTPSKLK 2519
 Db 2211 THQISRPNPNBPGFVUNSOQRQYEMQLQETQQLQMOOKYLEBIDG----- 2258
 QY 2520 VELIQOVNKHNAVIEHLKQKSKMTPAERENORMIVCNQVYKYLKIDKEEKOAKK 2579
 Db 2259 -----AHRKSKKALSAKOR-----TAKK 2276
 QY 2580 RKRESEVQKRSQONATKLSALLFKKEDLRAELKKRALLDKLOLEVOEELKRLDK 2639
 Db 2277 AGREPEDEAEOLKHTEQOSVQKQLEDIRK-----QCKEHAELIEDYRLK 2323
 QY 2640 KEDKMLQAOATVAAP-CPVTPVLPAPAPAPPSPPPGV-----QHTGLSTPTL 2691
 Db 2324 QOQ---QCAMAPPTMPSVQOPRPLIGATPTMSQPTPRMVPQOLQHOHTTVISGHTS 2380
 QY 2692 PV 2693

Db 2381 PV 2382

RESULT 7
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 108
; LENGTH: 1367
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 2.7%; Score 401.5; DB 10; Length 1367;
Best Local Similarity 21.2%; Pred. No. 2.8e-10;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;

QY 1261 DEEG---KIGCSE-----SNSTLENSDPYVSIQDSSEDMIVONSSESISQEFRT 1308
DB 191 DEPGFWYMNDCNNGCGTKSTTTSSTSSSTTSSTSS-----TTTSSTSSST 243
QY 1309 RQDVEVLEPLKCELVSGSTGNC--EDRLPVKGTETANKKPSQOKKLEDERPVNKKSSDI 1366
DB 244 TSTSESSSTSSSTTAPATPTTSTCKEKPPTPTTSTCKEKP-----PHHDTTPT 296
QY 1367 KLNKTTDKNNENRESEKKGORTSTFOINGKDNKPKIYLKGECLKESRVSQVSGNEPR 1426
DB 297 KKKTTTCKTKKTTTPVTPSSSTTE---SSAPVTPSSSTTSSSAPVTSSTTE-- 350
QY 1427 VNNINKIIPENDIKSLTVESAIRPINDVIMEDFENRNSSEYSHLLSSDAEGNYRD 1486
DB 351 --SSSAPV--TPSSSTSSSAPVTS-----STSSSAPVTSSTSSSAP 394
QY 1487 SLETLPTSTESDS---TOTTPSASCPESNSVNOVEDMEIEFSEVKKVTSPTSE---- 1539
DB 395 VTPSSSTSSSAPVTSSTSSSAPVTSSTSSSAPV--TSTSSSAPVTSSTSS 453
QY 1540 -----EESN--LSNDFIDENGLPINKNNVNGESKRKVTITEVTMTST--VA 1583
DB 454 SSAPVTPSSSTSSSAPVTSSTSSSAPVTPSSSTSSSAPVTSSTSSSAPV 513
QY 1584 TSKVIVIKVEKDKOTVVSSTENCASVTITVTYTKLSTPSTGSGVDIISVKEOSKTV 1643
DB 514 TPSSSTSSSAPVTPSSSTSSSAPVTSSTSSSAPVTPSS-----STSSSAP 568
QY 1644 VTTVTDSLITTTGGTLVSMYTSKE--YSTRKVKLMKFSRKKTPSSGALSYKRFYRK 1701
DB 569 VTSSTSSSAPVTPSSSTSSSAPVTPSSSTSSSAPVTPSSSTSSSAPVTS 628

QY 1702 STKKSIFVLPNDDLKLIARKGIREVPYFNNAKPALDIWPPSPPTFGITWRXRLQTV 1761
DB 629 STSSSSA-----PVTPSSS-----TTE 647
QY 1762 KSLAGVSLMLRLIMASLRDDMAKVPFGGSGTRTETSETITT-----TEIR 1811
DB 648 SSAPVTPP-----SSSTSSSAPVTPSSSP--TSSSAPVTSSTSSSAPVTSSTTE 701
QY 1812 RDVGPYGINFEYCIKRIIOPIGV-----EIPKEPPTQGRGLRBSAL----- 1854
DB 702 SSAPVTPSSSTSSSAPVTPSSSTSSSAPVTPSSSTSSSAPVTSSTSS 761
QY 1855 ---RKRRETPKOTGPVITETWAELELWEIRAFAREREKKAQAAVEQAKRLEQ 1910
DB 762 APVTPSSSTSSSAPVTPSSSTSSSAPVTPSSSTSSSAPVTPSSSNTSS 821
QY 1911 KP--VIANTSTS-----PSSSTTSTSPAKKVAVAISGSVTTGTMVLTXYGS-- 1959
DB 822 APSSPSSSTSSSVPTPSSSTSSS-----APVSSSTSSSAPVTPSSSSN 874
QY 1960 -----PATYTPQONKHFQEPATWVKQOGNSGVQVQCKVLG-----TIIS- 2001
DB 875 TSSAPSSIPFSSTTESST--GTVTPSSSKYIPGSOTLESVSSTTTTIVPTKTTSVTT 933
QY 2002 -----STGTSQ--QFTSPQRTATVYTRPNTSGSGTNSNOVITGPQIRPM 2048
DB 934 PSTTTTTCVGTGNSAGETTSKSPKVTYTV--PTTTTSTVTSSTTTT-----T 985
QY 2049 TYIRPVLQ---OSTLG---KAIIRT--PVAVQCAPQVWTOIIRQOPVSTANS-----AP 2096
DB 986 TVCSGTNSAGETTSKSPKVTYTVPCSTSPSEASESTTSPTTPVTVTVTVTE 1045
QY 2097 NTVSSTPGOKSLTSANSTNIOSS--ASOPPRQOGV--KLTMALQTLQFGHNOGLT 2153
DB 1046 YSISTKPGGE--ITTRVTVNIPPTTYLTTAIPPSVTVTVNPTTTTTCSTGTRNSA-- 1102
QY 2154 VVIQGGQTTGLOLIPQGVTVL---PGPQOQLOAMPNGVORFLPTPLATATPAS 2209
DB 1103 -----GERTSGS--PKVTVTVTPCSTGTEYTERA-----TTLVTVTA-- 1138
QY 2210 TTTTIVSTTAAGTGEOROSKLSPOMOVHODKTLPPAOSSSVGPARKQPOTAOPSARPOQ 2269
DB 1139 -VTVTVTVTSSGTNSAKKTTTG--YTTKSVPTTVTVTVLAPS----- 1178
QY 2270 TQPSAPQEPVQTPQVQGTVYSSHVPSEAOPTHQSSKPOVAQOSQOPSNVQOSPV 2329
DB 1179 ----APVTPATNAVPTTTTTECSATNAAGETTSKATTVSSASAGENTAPSAT-- 1231
QY 2330 VQSPQTRIRPST--PSQLSPGOQSOVQTT--TSQPIPIQPTSLQIPSOQPOSOPOV 2384
DB 1232 --TPVTTAI--PTTVITTESVGTNSAGETTTGTYTTSIPTTYITTL--IGSNGAKNYETV 1287
QY 2385 QSSTQTLSSGQTLNOVSVSPSRPQLOIOQPOPOVIAYVQLOQOVLSQISQVAYQIO 2444
DB 1288 AATV-----NPISIKTTS-----OLATTASASVAPV 1315
QY 2445 AQOSGVQOIKLOLPQIOQSSAVQTHOIONV--VYQOAS 2483
DB 1316 TSBS-----LTGPIQASGSAVATYVPSISSTYQGA 1348

RESULT 8
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US2002076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodges, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.

Qy	1885	RAFAERFEKEKAQAVEQAKKRLKEQOKPVIATSTSTSPS---	STSTISIP-----	AQK	1935
Db	3772	-----	PPPPPGTGTPTPTPTTTTTVVVPPPPPGTGT	3804	
Qy	1936	VMAVAPISGVT-----	TGKWLTTTKGSPAVTVFOONKNFHOATFVWVKOGOSNGV	1988	
Db	3805	PTTPTPTTTTTVPPPTPTGTGTPTTTPTTTTTVVPPTPTGTGTPTTTTITTTT---	3861		
Qy	1989	VQVOQKVLGIIIPSSGTSGOQFTSFQPKRAVTVIRPNISGSGGTTNSQVITFQPIRPGM	2048		
Db	3862	-----	VTPTPTPTGTGTPTT-TPTTTTVVVPTPTPTGTGTPTPTTPTT---	3908	
Qy	2049	TVIRTPVLOSSTLCKAIRPVMVWQPCAPQOVMTOIIRQGPVSTASANTVSSTPGOKSL	2108		
Db	3909	TPPTPTGTGTPTPTPTTTTTVTP-TPPTGTQ---	TPPTPTTTTTVVVPPPTPTGT	3963	
Qy	2109	TSATSTGNISASAP-PRPQOGVNLVMAQLTOLGCHGNGLTVVIGOGGHTGOLQ	2167		
Db	3964	QTPTTTTITTTTTVTPPTPTGTGTPTTTTITTTT-----	VTPTPTPTGTGTPTTT	4015	
Qy	2168	LIPGCVTLPGPQOQLMQAMPNGVORELFPPLAT	-----	ATTASTTT	2212
Db	4016	PITTTTTTPPTP-----	TPGT-QTPTPTPTTTTTVPTPTPTGTGTPTTTPTTT	4066	
Qy	2213	TTVSTTAAGGEGEQRQKSLSPQMVHODKLLPRAOSSVGRAPAAQGTQAPSKRPQOP	2272		
Db	4067	TTVPTPTPTGTGTPT-----	TPPTTTTTVVVPTPTPTGTGTPTTTPTTTT	4114	
Qy	2273	QSPAPQEVQOPEVQOQTVSSHVPSQAPPTNASSKRPQVMAQSOQSNVQOGSPVRVQS	2332		
Db	4115	VTPTPTPTGTGTPTTTPTTTTTVTPPTPTGTGT-PTTTTITTTTTVTPPTPTGTGT	4172		
Qy	2333	PSQTRIRPS---	TPSOLSPQOSQOVTTTSQPTPIQPHSTLQIPSQGQPQ-----	SOPOV	2384
Db	4173	PTTPTPTTTTTVTPPTPTGTGTGPPHTHSTAPIALTLTNSNPPSSPTQSRSTSPLT	4232		
Qy	2385	QSSQTQLSSQOTLQNVSSPSNP	2408		
Db	4233	ESTTLSTLPPELMTSTAPPSTP	4256		
RESULT 10					
US-09-815-242-5816					
Sequence 5816, Application US/09815242					
Patent No. US20020061569A1					
GENERAL INFORMATION:					
APPLICANT: Haselbeck, Robert					
APPLICANT: Ohlsen, Karl L.					
APPLICANT: Zyskind, Judith W.					
APPLICANT: Wall, Daniel					
APPLICANT: Trawick, John D.					
APPLICANT: Carr, Grant J.					
APPLICANT: Yamamoto, Robert T.					
APPLICANT: Xu, H. Howard					
TITLE OR INVENTION: Identification of Essential Genes in					
TITLE OR INVENTION: Prokaryotes					
FILE REFERENCE: ELITRA.011A					
CURRENT APPLICATION NUMBER: US/09/815,242					
CURRENT FILING DATE: 2001-03-21					
PRIOR APPLICATION NUMBER: 60/191,078					
PRIOR FILING DATE: 2000-03-21					
PRIOR APPLICATION NUMBER: 60/206,848					
PRIOR FILING DATE: 2000-05-23					
PRIOR APPLICATION NUMBER: 60/207,727					
PRIOR FILING DATE: 2000-05-26					
PRIOR APPLICATION NUMBER: 60/242,578					
PRIOR FILING DATE: 2000-10-23					
PRIOR APPLICATION NUMBER: 60/253,625					
PRIOR FILING DATE: 2000-11-27					
PRIOR APPLICATION NUMBER: 60/257,931					
PRIOR FILING DATE: 2000-12-22					
PRIOR APPLICATION NUMBER: 60/269,308					
PRIOR FILING DATE: 2001-02-16					

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Ver
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

```

Query Match	2.6%;	Score 389.5;	DB 10;	Length 2478;
Best Local Similarity	17.8%;	Pred. No. 1.8e-09;		
Matches 447;	Conservative 354;	Mismatches 1008;	Indels 699;	Gaps 105

Qy	475	DPEQKSEEPTEVGDKNUSVANSILGDTNTATSEETSPBGRSGJSEFPDSSMMXK	534
Db	40	EONOPAONOPADANTOPNANAGAO-ANPTAOPAPANOGOPA---VOPANOGGONAP	95
Qy	535	KVASELPODVP-----EPEPKTCESSANTTATSTSIOPULEN-----SNSSELSOSSE	583
Db	96	AGGAAGOPNTOPAGGONADPNNAQAOPGONQATPPANAOAGGONNOATPPNNATPPANOTPA	155
Qy	584	SAKADDPENGERSHETTPVSIQOEIYQDFTSEKSTGELSESPQAGKASGSTRITITRLN	643
Db	156	NAPAAOAPA-----APVANAQA-----TODPMASTGEGSINTTLTFDD	194
Qy	644	PDKLSOLQKSQVAAAHAHEAN-klPEKEGEVLVYVSGOESISLSTKKREVIMKGINNYF	701
Db	195	PAISTDERODPTVTVDKNGVSGSLNNK-IEFVNSE-----LRSRMPDKNPNQ--	246
Qy	702	KLQGEKRYVHNQYSTNSPALNKHQREHDKRRHLAKHFCLLTPAGEFKMGVSHG--SK	760
Db	247	-----QAKRVAALGRVANABSTD-----HGNNGSK	274
Qy	761	VLIT--STLTLITTOLENNIPSSFFHPNASHRANWIKAYOMCSKPREFALALILECA	817
Db	275	TYNKPSSELLINTFTMQTSKOG-----ATNLVYIDAKKNTLATAVVA	319
Qy	818	VKPYVMPLEPMEFELGHTRLHRM-TSIREKEKVKKKKKQOEELMOQATVWY-----	871
Db	320	K-----TGTAHLKRVLPDADARLDLOLIPNTAVADAIRITTKDKGYCTSYFD	367
Qy	872	---TFPVHQWQK-----GEEVRYT---GYGWSMISKTIVYRFVFKLP-GVTN	915
Db	368	NVGLFSGSHLYVKRRDLAPKATINKETTITELGNNNGNGASLSKAQOFXYETLLQGYTY	427
Qy	916	VNYKRSLEGT-----KNNDEMDESDKRKCSRSP-----KKIKIEPD--	953
Db	428	VN--NSLTTTPNGNEDSTVLKMNVTYDQNAKVFYTSQGVTTARGTHTKEVLPPDKSL	485
Qy	954	--SEKDEYKSDAKGADQONEMDISKTTEKKODDYKELLDSDSKCKEPEHEVDDMKT	1011
Db	486	KLSYKVVYANIDPKNIDFNE---KLTYRTASDV-VINNAQPEVYTLTADPESVAVENNK	540
Qy	1012	ES---HYNCOESS-----OVDVNVNSEGHILRTSYKKRTSKSKLDGLE	1052
Db	541	DALOOQVANSQVDNSHYTTASIAEYNNKLKQOADITLMEADANHAKTA--NRASQADIDGLV-	597
Qy	1053	RRIKOFTLEEKORLEKIKLEGIGIKGKSTJNSNKSILSESVP---TTAKKEGQOSDSMKO	1105
Db	598	TKLQALIDNQAALAIELDTMAQEK---VTAQOSKSKVTQDEVAALVTK-----	642
Qy	1110	EOSPANANDOPEDLIQGCOSQSDSVLEMSDPSTTKN-----LYPKDRLDVYSIRSEPT	1164
Db	643	-----INNKNNAIAELNKOJTAAQV-----TEKONGIAVLEQDVTTPV---KPOA	687
Qy	1165	KCPKONSTENDIEKVSDDLASROEPTKSKTKNDFFIDSKLASADIDGTLICNKKPL	1224
Db	688	K-----ODIQAVTTRKOQIKKSNA5-----LODEKDVANDKIGKLETATAIDI	731
Qy	1225	IQEESDTIVSSKSLASHSVPKSTND-RDTPPLSRAMDEFGKIGCDSSEJNTLENSDPV	1283
Db	732	DAATTNAAVEAIKT-----KAINIDINQTTPTTA-----KAAALEEFDEV	772
Qy	1284	SIDDSSEDMIVONSNEJSEOPTRBODEVLEPIKCELVSQ---ESTGNCEDRLPVK	1339

QY 2386 SSTQTLSGGOTLNQVSSPSRPOLOIQOPQVIAVPOLOQVOY-----L 2432
Db 1766 TITLYAYADOKNNNISADT NATODEKQO-----AIKQYDVQVIALESINNNGVNDGY 1818
QY 2433 SOIQSVVAVQIOAGSGVPOQIKLOLPQIOQSSAVQTHQVNVVQVQASVQEOLOQRYQ 2492
Db 1819 DDAITGQKALIDAIQVDAVQPRKANOAEVKADPTKESIDQSDQLAEERT--EALAMIK 1876
QY 2493 QLFDOQ-----QKKQOQOIEINVTSPSKLILKVELIQVQVKNHVAIEHLK 2539
Db 1877 QITDQAKQGITDATTATTAWEKAKAQGLEAFDN-----IQIDSTEQO-----KALIEELE 1924
QY 2540 OKKSMTPAEREENQRMIVCNOQVMIILDKID-----KEEKOQ-----AKKKR 2582
Db 1925 T-----ALDQIEAGVNVNADATTEKAEFTALDILSKAT 1960
QY 2583 ESEVEOKRSKONAT-KLSALLEFKHKEOLRAELI-----KKRAL-----IDKQLOIEVOEE 2631
Db 1961 EDISDQTTNAETVKNASAL-----EOLKQORINPEVKKNALEIREVVK--QIEIIRK 2013
QY 2632 LKRDLIK-----KEKDMQLOAQATVAVAPCPPTVTVLAPRAPPSPPPRG 2679
Db 2014 ADADASAKETARTDLGRYDFRADKIDKQOTNAEVALQNT--IPATEIIVQNDPDAN 2071
QY 2680 VQHTGLSTPTLPVASQKRKEEEKDSSSKKKMIITSKETKKDT 2727
Db 2072 DTNNGI-----DNDATANSNMNATPENQPNVSET 2103

RESULT 12

US-08-681-219-32

; Sequence 32, Application US/08681219
; Patent No. US20020058607A1

; GENERAL INFORMATION:

; APPLICANT: Takaaki Sato and Junn Yanagisawa

; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN

; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF

; NUMBER OF SEQUENCES: (Pdz/DHR) DOMAIN AND USES THEREOF 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/681,219

; FILING DATE: 22-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JKN

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2843 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-681-219-32

Query Match 2.4%; Score 361.5; DB 8; Length 2843;
Best Local Similarity 17.3%; Pred. No. 3.e-08;

Matches 412; Conservative 325; Mismatches 846; Indels 795; Gaps 100;
QY 373 WEAELEKILEEMEEIHRHDDITEDLTNKRAGSKNFLLAANEELIESIRAKGDDDNK 432
Db 699 WDAVAVSMKLNLIHSKHKM-----AMGS-----AALRNLMANRPAAKAKDA-NIM 743
QY 433 SPE-----ETEKDK---NET-----ENDSKDEKNREPEFDOSLEKD----- 466
Db 744 SPQSSLSPLHYRKQKALEAELDAQHLETFDNDINISPRASHSKORHKOQLXGDIYFDT 803
QY 467 --SDDKTPDD-----DPEQKSEEPTEVGDKGNSVANIGDNTNATSEETSPSEG 515
Db 804 NRHDDNRSDNFNTGNMTVLSPYLTNTVLTPSSSSRSLDSSR-----SEKDRSLER 854
QY 516 RSPVGC-----LSETPDSSMAEKKVASLEPQ--DVPEEN--KTCGSSNTATSTSI-- 564
Db 855 ERGIGLGNHNPATENPTGSSKRLQISTAAQAKWEEVSAIHTQEDRSSSTTELIC 914
QY 565 ---QPNLENSNSSSELNSSQSESAKADDPENGERESHPTVSTOIEIVGDFTEKSTGEL 621
Db 915 VYDERNALRRSSAAHTHSNTYNTKS---ENNRKCSMPYA--KLEVKRSSNDLSNV 967
QY 622 SESPGAGKAGSGSTRITTRLNPDSKLSQLSK--QQVAAAHEANKL-FKEGKEVLVN- 677
Db 968 SSSDGYGKRGQMKRPSIESYSEDEDEKFCGYQYPADLAHRIHSANHMDDNDGELDTPINY 1027
QY 678 -----SQGEISRLSTKKEVIMKGINNNYKLOEGKRYRHHOYSTNSPALKHQH 728
Db 1028 SLKYSDBQLNSGRQSPSONERMAPKHIIEDETKOSBORO--NOSTT--IPVTTEST 1082
QY 729 REDHDK-RRHLAKFCLTPAGEFKMNGSVHGSKVLTITLRLITOLENNIPSPFFHPNW 787
Db 1083 DDKHLKFPQHFQGGQECVSPYRSRGANGS-----ETNRVGS----- 1117
QY 788 ASHRAMWITAYQMSKPREPALLALIECAVQVYVMLPIYREFLGTRILHMTSIRREK 847
Db 1118 -NNGIQNNVSQSLQED--
QY 848 EKVKKKEKKOEEETMQOATWYTFEPVKHQVWKQGEERYVYGGMWSIKTHYREV 907
Db 1139 KPTNYSERYSEEF-----QHBEERPTNY----- 1162
QY 908 KPLPGNTNVRKRSLEGTKNMDEMDES-----DKRCKSRPKKIKI 950
Db 1163 -----SIKYNEE---KRVHDPIDYSLIKATDIPSSQKOSFPSKSSGQSKTEHM 1211
QY 951 EPQSEKDEYVGSAAKGAQDNEMDISKITTEKKODVKELDNS-----DKP 997
Db 1212 SSSSEMTSPSSNAKR--ONOLHPSSAOSRSGQPKAATCKVSSINQETIQTVCVEDTP 1268
QY 998 -CKEEMEVDDDMKTESHVQCESSQ-----VDVVNVSEGFHLRT-----SY 1038
Db 1269 ICSRSRSLSSLSASADEIGCNOTQEOADSANTLOJAIKEIKGTGSAEDPEYVAVSQ 1328
QY 1039 KKKTKSKLDGLLEIRIKOFTLEKORLEKIKLEGKIGIKGTSTNNSKN-----LSESP 1093
Db 1329 HPRTKSSRLQG-----SSLSESARHKAVFESSAKSPSKGAGTAPKSPREHYVQETP 1381
QY 1094 VI-TKAKGQSDSMQEOQSPNANNQOPELILQCGQSDQSVLRMSDPTHTTKLTPKXR 1152
Db 1382 LMFSCRQTSVSSLDSPFSRSIASSVQSEP-----CGGWVSGIISPDLDPSPQOTPPSP 1435
QY 1153 VLDDVSLRSPETKCPKONSTIENDIEKVSQDLASRGQEPKSKTKGNDFFIDSKLASAD- 1211
Db 1436 -----SKTPPPPPQTQTKRE-----YKKNKAPTRAEKESG-----PKQAAVNA 1474
QY 1212 -----DIGTLI-----CKNK-----KPLQOESDITV-----SSSK 1237
Db 1475 AVQGVQVLPADQTLHLHATSTPTDGRSCSSLSALSLDEPFQKQVDELRIIMPVQGNNG 1534
QY 1238 SALHSSVPKSTNDRDATTPLSRAMDFEGKLQCGSESNTLENSDTYIDQSSSEDMATYON 1297
Db 1535 NETESQPKESNENQ-----EKEAEKTIIDSEKD---LDDSDDD----- 1570

```
OY 1298 SNEISBQFRTREOVLEVLKCELVSGESTGNCEDRLPVKGTANGKPPSOQKLEER 1357
Db 1571 -----DIEILE--EC--IISA-----MPTKSSR-KAKKPAOTASKLEPP 1603
OY 1358 PVNKCSDOI---KLANITDDKNNENRESEKKGQRTSTQIINKDKPKLYLKGCECKEIS 1414
Db 1604 PVARKPSQIPYKLLPSONRLOPOKHVSFTPG-----DDMKPVY---CV----- 1644
OY 1415 BSRVSGVNEPVKNINIKIIPENDIKSLTVKESAIRPFIINGDVIMEDFENRNSSETKSHL 1474
Db 1645 BGTPIINFATNSLSDLTLESPPNEL-----AAGEGVRGGAQSGEPFKRTIPIEGR- 1695
OY 1475 LSSDAEGNYRDSLETLESSTKES-----DSTQTT 1503
Db 1696 -STDAOGKGTSSV--TIPELDONKAEEDILAECINSAMPKGSHPFRVKKIMDOVOQA 1753
OY 1504 TPSACPSNSNVQVEDMEIEISEVKKYTS-----SPITSE 1539
Db 1754 SASSSAPKKNQLDGKK--KKTPSPVKPIPONTREYTRVKNADSKNNLNLAERVFDNKS 1811
OY 1540 BERNLSNDFIDENGLPIKNENVNGE----- 1565
Db 1812 KKNLNKNSKDPNDKLPNNEDRVGRSFAFDSPHHTPIEGTPYCSRNDLSLDFDDDD 1871
OY 1566 ---SKRKVTIEVTMTSTVATESKTVIKVEKDKQTVVSTENCASKSTVT----- 1613
Db 1872 VDLSSREK---AELRAKKEKESKAYTSHTELTSNQOSANKQALAKQINGQPKPILO 1928
OY 1614 -TTTTTIVKLSPTSGSGSDIISVKEQSKTVTTVTDSLTGTGTLVTSMTVSEYSTR 1672
Db 1929 KOSTPQOSKIDIPDGAATD---EKLONFALIENTPVCFS--HNSLSLSLSDIDENNNK 1982
OY 1673 DKVKLMKFSRP-----KTRSGTALPSYRK-----FVTKSTKKSIFVNDL- 1715
Db 1983 ENERPKEETPEPOSOEPRKPOASGAPKSFHVEDTPVCFSRNSSLSSLSDEDDLQBC 2042
OY 1716 -----KTLARKGIREVRYFNANRP-----ALDIWPPSPRPRTFGIT-- 1753
Db 2043 ISSAMPKKKKPSRLKGDNE-----KHSPRNMGILGEDLTLKDIOQPDSEHGLSPDS 2096
OY 1754 ---RRYRLQTVKSLAGVSLMLRLMASLRMDMAKVPBGSGSTRTESEIEITTELI 1809
Db 2097 ENFDKKAIOEGANSTIV-----SSLIHQAAAACL-----SROASDSD- 2136
OY 1810 KRWDGPIREYECIRKICPIGVETPEKERTPOQRKGLRSALRPKPPETPKGTGPIV 1869
Db 2137 SLKS-----G-----SLGSPHLTPQOEEKFTSNKGR--ILKPGKST- 2175
OY 1870 IETWAEDE-----LELMEIRAFARVEKEKAQAVEQAKKRLDQKP- 1912
Db 2176 LETKIESESKIGKGGKVKYKSLITGVKYSN- 1912
OY 1913 --TVIATSTSTSTSTISPAQKVMAPIGSAVTGTCKMVLITKVSAPATVTOQKN 1970
Db 2227 GRTMHIPGVNRSSSTSPVSKKGPLKTPARKSPSEGTATTSBRGAKPSV- 2278
OY 1971 FHQFATWAKOGOSNGVQVQOKVILGIIIPSTGSOQFTSFQORTAVTIRPNTSGSG 2030
Db 2279 -----KSELSPYARQTSQI-----GSSK-----APRSRSGR 2305
OY 2031 GTTNSQVITGPQIRPGMTVIRTPLOQSTGLKAILRTPVWQPG---APQOVMTQING 2086
Db 2306 DSTPS-----RPAOQPLSRPIOSGRNS-----ISPRNGNISPPKISQLPRT 2348
OY 2087 QPVSTAASAPNTVSTPGQSLTS-----ATSTNIOSSASQPPRPOQGVKYLMA 2137
Db 2349 SSPSTA---STRKSSGSKMSTYSPGRMSOONLTKQTLKSNKASIPRSESAS---K 2399
OY 2138 QLTOLTOGHGNGGLTVVIOGQOTTGOLQ-----LIPQGVTLPGGQOLMAAMPN 2190
Db 2400 GLNOMNNGANKKVELSRMSSTKSSGSESDNSERPVLVKROSTFIKEADSPILRRKLEES 2459
```

```
OY 2191 GTVORFLPTPLATATATTTTIVSTTAAGTGEOROSKLSF-----QMOVHD----- 2239
Db 2460 AS-----FESLSPSSSRPAPSPRSQAOTPV-----LSLSPLDMSLSTSSVQAGW 2504
OY 2240 KTLPPAOSVGPAPAKAOPQOTAPASAPPOQOTOPSPAPQEVOTQOTVSSHVPS 2299
Db 2505 KRLPPLSLPTTIEVNGRPAKRDIAHSSESRLPINSQGWKRE---HSHSSSLP 2561
OY 2300 AOPTHAOSKPOVAOQOSPOSNVQOSPVVQSPQTRIRPSTPSOLS--PGQOSOVQTT 2358
Db 2562 SWMRRTGSSSSILSASSESSEKASDEKHVNSISOT--KQSKENQVSAKGWIRKIKENE 2619
OY 2359 SOPITPIQPTSLQIPSOQPOSOPOVOSTOTLSQOT 2356
Db 2620 FSP-----TNSTSQTVVSGAT 2635
```

```
RESULT 13
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSHOT, Astrid Adriana Anna Maria
; APPLICANT: KOHN, Jennifer Leigh
; APPLICANT: WEISS, Bertlram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 472505
; CURRENT APPLICATION NUMBER: US/09-764, 176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the
; OTHER INFORMATION: Ictc acid sequence of AAP-
US-09-764-176-7
```

```
Query Match 2.4%; Score 354.5; DB 10; Length 1400;
Best Local Similarity 18.2%; Pred. No. 3.1e-08;
Matches 290; Conservative 240; Mismatches 584; Indels 481; Gaps 56;

OY 274 HLECYKPLIEEYED- 274 -----QCEVCAHVPGTDCVALELOKKKPIRRHPGY 324
Db 61 HLEKMKIKGVTADRWKERYLIKYLCEQFDNLKFKNTIN-----EEDADTMRLOPIGR 115
OY 325 DSRRRKYMPL-----NRLLIEEDFENENEKIKWYSTKVQALAILDCLDKDYWEALCK 379
Db 116 DKDGLMYWQLODDHNVRYIEQDDQ----- 143
OY 380 ILEMKREIIRHMDITEDLTNRKGSNNKSLAANEILSIRAKKGDIDNVKSPBETEK 439
Db 144 -----GSMKCIVRNNEIETLALILKAOIDPVLKNSQO 179
OY 440 DKNETENDSKDAEKNRBEFEDOSLEKSDDKTPDDDPQCGKSEEPTEPVQDKNSVANSIG 499
Db 180 DMSREPSLEDEETKKEETPKOEOKES-----EKMKSE- 230
OY 500 DNTVATSEETSPSGRSPVGCLETPOSSMAEKKVASELPQDVPPEEPNKTCESSNTSA 559
Db 231 EETTVKKEK-----DEKELVKLPIVLYKLEKPLPENDEKIIIEBDS 273
OY 560 TTTTSIQNLNENSSSELNSSQSESAKAADPENGERESHFVSIQIEIIVGDTFSEKSTG 619
Db 274 FKENVK-----IKVEVKECRADPKDKRSSMEKPVVAQEPRIEFGNKKSSH 320
OY 620 ELSESPGACGKSGSTRITTRLRNPDSKLSQSOVAAAHAKKLPREGKEVLYVNSQ 679
Db 321 EITER-----STEETKLNDOAKIPLKKREIKLSDDFDPSV--KGPLCKSVYPT 369
```

OY	660	GEISRLSKKKVILKGNINNYFLGEOGKRVYHNGYUNSNALKKHQHREHOKRRLLA	733
Db	370	KEFLKDEIKOEETCKRITITTLAHEGK-OLVNGEVSERYAPN-----FTPEIE	420
OY	740	HKFCUTPAGE-----KNGSVHSGKVLTL--ISTRLTITOTLENNIPSSFFPHN	786
Db	421	TKFYETKEESYPSKDRNITTEGNGEINSVITSMKGTGELKEPAJLRKADSS-----	475
OY	787	WASHIRAWIKAYOMCSR-----PREPALALILECAVVKVYVMLPIWREPLHTLRLH	838
Db	476	-----ISVLEHSHKAOIEBDDPEMETSLDSSEMA-----KDLSSKTAALS	517
OY	839	MTS---IEREEKYKKKKKOEHEETMOQATWTKYTF-----PYKHVWKOGXEY	887
Db	518	TESTCMGEEKSPKTKDKRPPLIECLEKEKSKKFFLDKQAKRLSPIREVPKSTLESE	577
OY	888	RVTYIGGWSMISKTHVRYEVPKLPNGTNV-----NYRKSLEGTKN	927
Db	578	KPGSPEAETSPNNIDHCEKELASEREVEECOSTYVGOGSVKKVDTLETLKEDEEFTKY	637
OY	928	NMDSNMDES-----KRCSSRPKKIKIREPSEKDEVKSGDAKGAONMDISK	977
Db	638	EMD-NLDNAOTSGLEEPSFKSGMOKSKRYKALVPEEIT-----ASEN---TE	682
OY	978	IT-EKKDQDYKELLSDSDKPCKEEPMEYDDDKTESHVNCOSSQVDVYVNSEGFLT	1036
Db	683	ITSEROKEGKIKLIRIISRKKKPPDSPPKV---LEPE---NKOEKTEKEEKTNVGRTLRL	736
OY	1037	SYKKKTSKMLDGLTER-----IKOFTLEKORLEKI	1065
Db	737	SPRISSRPATAVAETIRODKADKKRGEDEVEESTLAOKTKKEIKULSEKDTNSKV	796
OY	1070	KLEGGIKIG-----KTSJNSKNLSESPVITKAKECQSDSMKRODOSPANNDDP-	1120
Db	797	KPKKGVWMTGSRTGRWKVYSNDESGSGSEKSSAASE--EEBEKSEFALILADDERCK	854
OY	1121	-----EDLIOGCSQSDSSVLMSPSHTKNLP-----K	1150
Db	855	KCGLPNHPEILLCDSDSGY-----TACLAPRLMIIPDGEWFCRPOCHKLCEKLE	907
OY	1151	DVRLD-VPSIRSPETKCPKN-----SIENDIEKVSIDLASROEPTK--SKTKGN-	1198
Db	908	BQLODDLVALKKKEARERKRELRVYVIGISTENIIPQEDRPSDEDEBKKSOKSKANLL	967
OY	1199	-----DEFIDS---KLASAD-----DIGTLCKKKRP---LI	1225
Db	968	ERRSTRTRKCIYSRFEFDEAIDEAIEDIIEKDEGGVGRGKDITIGHRKDITSLD	1027
OY	1226	QESDITVSSKSKSLASSVYKSNDDNDARPLSRANDBEKLGCSDESSTLENSDTYSI	1285
Db	1028	EEKENKRRPRAAARAKKRRRLND-----LDSDSMLDEESESDEFKI	1070
OY	1286	QDSSEEDMIYONSNESISEQFTRIEDOVLEPLKCELVGSESTONCDBRLPVKGTENG	1345
Db	1071	SDGSODEFVYVSDENPDESEDPSND-----SDTFCGRLL---RRHP	1111
OY	1346	KKPSQOKKLEER--PVNKGSDQIKLKNITDKKNNR-RESF-----KKG	1386
Db	1112	SRPROSRRRLRRTPKKYSYD-----DDEEESERKSRSDSFSDFSDFETVETRRRS	1166
OY	1387	QRTSTPQINGK-----DNKRIYLGCEKKEISESVVSGVNGEYKPVNNINKITIPENDKS	1441
Db	1167	RRNQKRINYNKDESEDSGSKSLRKKELEIRVHKRRLLSSSEEBEYLSKNS---EDDEPLA	1223
OY	1442	LTVKESAIRPFIINGDIVIMEDFENRNSSETKSHLLSSDAEGNYRDSLTLPSJTKESDTQ	1501
Db	1224	KESKRYSRKRGSRJSTDESEADEEEEBEKGPS-----RKRLHRI-ETDEEESC	1271
OY	1502	TTTTPSACSPSNSVNGVYEDMETITSYKVKVYSSP--IJSBEESNLNDFIDENGLPIKN	1559
Db	1272	NAHQDAQPARDSQPRVLSEDS-----TKPPIRISDEEEDEN--VKVGSPLDYS	1323

```

QY 1560 --ENVNGESKRKYITTEVTMTSTVAAEKSTYIVKEAGDQTYVSSNENAKSVITTT 1615
                                     : : : : : : : : : : : : : : : : : :
Db 1324 LVDPSTSTGSGPQKAL-----ENLIGKPTKSPKDN 1357
                                     : : : : : : : : : : : : : : : : : :
QY 1616 TTTVTKLSTPSTGSDVSDISVKEQSKTVTTVTVD 1650
   | : | : | : | : | : | : | : | : | : | : |
Db 1358 TASAISLASNGTSGGQACAPPEEEDDLAVTDLVD 1392

RESULT 14
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match 2.3%; Score 342; DB 10; Length 2368;
Best Local Similarity 17.6%; Pred. No.1,9e-07;
Matches 436; Conservative 342; Mismatches 866; Indels 838; Gaps 111.

QY 475 DEQCKSEEPTEVGVGDKGNSVSNALGDNNTNATSEETSPSEGRSPVGCISETPDSSNMAEK 534
   : : : : : : : : : : : : : : : : : :
Db 40 EGNQPAQNOQAPADANQPNANAGAQ-ANPTAQATPANQOQPA---AQPANQGOQANP 95
   : : : : : : : : : : : : : : : : : :
QY 535 KYASLEPDQV-----EPPNKTCCSSNTSATTTSTQPULEN-----SNSSSELNSSQSE 583
   : : : : : : : : : : : : : : : : : :
Db 96 AGGAQAPMTQPPAGGQNOADPNNAQAQAPQNOAAPNAQOQGNNOATPNNNATPANQOTOPA 155
   : : | : | : | : | : | : | : | : | : |
QY 584 SAKAADDPENGRESHTPVSIOEIVGDPETSEKSGELSESPGAGKAGASGTRITRLRN 643
   : | | : | : | : | : | : | : | : |
Db 156 NAPAQAQ-----ANNAQTQDPNASNTEGGSINTLTLPED 190
   : : : : : : : : : : : : : : : : : :
QY 644 PDSKLSQKSOQVAAAHEAN--KLFEKGKELYVNSQGETISRLSTKKEVINKNINNYF 701
   : : : : : : : : : | : | : | : | : |
Db 191 PAISTDENRQDPVTVTQKVNQYSLNNGK-IGEVNSE-----LRSDMFDKNNPQNV- 242
   : : : : : : : : : : : : : : : : : :
QY 702 KLQGGKTRVYHNQYSTSFALNKQHQHREDHDKRRHLAHKFLTLTGAQGEKKNAGSVHG-SK 760
   : : : : : : : : : : : : : : : : : :
Db 243 -----QARGVVAALGRVNAVNDSTD-----HGNGNGISK 270
   : : : : : : : : : : : : : : : : : :

```



```

Db      1938 T-1PA1EAIYFQNDPNANDTNSG 1959

RESULT 15
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12389
LENGTH: 2368
TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match      2.3%, Score 342, DB 10, Length 2368;
Best Local Similarity 17.6%; Pred. No. 1.9e-07;
Matches 436; Conservative 342; Mismatches 866; Indels 838; Gaps 111.

Qy      475 DEQCKSEBPTVGKNGSVSANILGDNNTNATSEETSPSEGRSPVGCUSEPTDSSMAEK 534
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      40 EONOPANOPAPADANTOPNANAGAO-ANPTAOPATPNANOGCPA---AOPANOGGAP 95
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      535 KYASELPQDVP-----EBPNKTCSSSNTSATTTSTIOPNLEN-----SNSSELNSSQSE 583
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      96 AGCAAOPTQAPGOGNOADPNNAQAOPNQAAAPANAOAGGNQATPNNNATPANOTQPA 155
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      584 SAKAADPENGERESTPTVSIQEIYGDPTSEKSTGELSESGAGGAGGSTRITRLRN 643
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      156 NNPAAAP-----AANAQODPNASNTGSGSINTTLTFDD 190
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      644 PPSKLSQLKSGQVAAAHAHEAN-KLFKEKEVLVNVSGEISRLSTKKEVINKGNINNYF 701
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      191 PALSTDENQODPTVYTDKVNKYSLLNNCK-IGFVNSE-----LRSMDPDKNNQNT- 242
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      702 KLGECKYRVVYNQVSTNSFALNKHOHREDHDKRRHLAHKFLTPAGEEKWNGSVHG-SK 760
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      243 -----QARGNVAALGRVANADST-----HGNFGISK 270
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      761 VLTl---STLRLLTQLENNIPSSFPHRWASHRAWIKAYQMGSKRREFALALALECA 817
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      271 TVNVPKDSLLINFTTMQTN-----SKO----- 293
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      818 VEPVWMLPTWREFLGHTRLHRTSTIEREKEKVKKKKKOEETMQQA-TWVKYTFPVK 876
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db	224	----	GATNLL-----	VIKDAKNTL	ELATVUNAK	GTALHFKVP	326		
Qy	877	HOVWK-----	-QKGEERYPT-----	GYGHSML-----	SKTHVY-----	RPVPLKPG	912		
Db	327	TDADRLDQFIPDNTAVADASRTITNKDGKYYSF	IDNVGLFSGSHLYKRNBDLAPKTN				386		
Qy	913	N-----	TNNYRKSLSEGT-----				930		
Db	387	NKEFTINTEIGNNONPFASLAKQOFXYEVL	PGCVIYVANSLLITTFPONGNEDSYLKNMT				446		
Qy	931	ENMEDSDRKCSNP-----	KKIKLEPD-----	SEKDEKXSDAKGADONENDIS			976		
Db	447	VNYDONANKVTFISQVTTARGTHTKEVLE	PDKSLKSLKYVWVAIDTPKIDFNE-----				502		
Qy	977	KITEKKODVVELLDSDS	DRCKBEPEYDODDKTSHNCOSSOVD-----				1024		
Db	503	KLITRTASDI-VINNAPPEVTLTADP	PSAVENTKKA-LQOOVNSQVDNSHTTASINEX				560		
Qy	1025	-----	VYNNSEGEHLTSYKKTKSSKLDGLERRIKOFTLEEKORLEKILEGII				1075		
Db	561	NKLKQADNLTILN-EDAHNVEFA-NRASA	AIDGLV-TKLQALILDNOAALIELDAKAOE				616		
Qy	1076	KGIGKTSJNNSKNLSSEP----	ITKAKECQSDSMQEOBSPANNDOEPDLIQGCSQSDS				1132		
Db	617	K-----	VTAAGQSKRTYQOEBAVALYK-----				645		
Qy	1133	SVLRKSDSHNTKNTLYPKDR-----	VLDVSI	RSPEKCPKONS	IENDIEEKVSDLASRQOE		1189		
Db	646	AIAEINKQTTAQGYTTEKDNGLAVLQD	-DYITPYVPAQKODI-----				697		
Qy	1190	PTKSKTKGNDEFIDDSKLASADITGILLIC	KNKPKPLIOESDITVSSKSAHLSVPKSTN				1249		
Db	698	IKKSNAS-----	LODEKDVANDKIGKLETKAIDDAATTNQOVEAKYT-----				745		
Qy	1250	DRDAT-PLSRAMDEPKLGDOSSE	NSLTENSPTSJODSEEDMIVONSNE	ISEGPT			1308		
Db	746	DINOTAPATTA-----	KAALIEFEDEVQAO-----				782		
Qy	1309	RODEVYLEPLKCLVSG-----	ESTGNCEDRLFPVKTEANGKRP	PSQOKLEERPYNKCS			1364		
Db	783	NEEVAELIERINAKVSGVAIEATTA	ODLEKRVKNEEIS-----				822		
Qy	1365	QIKLKNTDKNNNRESEKKGQTS	FOJNGDKNPKIYLIGECLEKTESR	VYVSGNVE			1424		
Db	823	-KIENTITDS-----	TOT-----				852		
Qy	1425	PKVNN-INKIIPENDIKSLV	KESATIPFINGDVI	MEDEERNSE	SETKSHLSBSAEGN		1483		
Db	853	ATVSNA	TNEEVALEAD--AAVEA	QOGHLDIOYVK--SKEVAD	TSKVLDIRINA---		903		
Qy	1484	YRDSLETP	PSIKEDSTOTTPPASCE	SVN-----QYEDMA	ELIESPVKATTSPI---		1536		
Db	904	-----	IOTAKVYK-----				950		
Qy	1537	-----	TSEENLSNDF--IDENIL-PIN-----				1576		
Db	951	KOEANTNIDAA	NTSDVTTAKDNCIA	INALNOYAATTKRSDAKAE	IAQKRSSEKTAIEAMN		10110		
Qy	1577	TMTSVATES	STVIAKEGDKOTVVSST--ENC	AKST-----VTTTTT	VTVKLSTPSTGG		1629		
Db	1011	DST-----	TEEOQAAK-DRVQDA	VYVTAADIDNAA	MTVDNAKTTEATIEA	IAI--TPDA--	1062		
Qy	1630	SVDIISV	GEOKTYVTTVYDLSLT	TGVLVTSMY	SKESTRDKYKLMKFS	RPKKSTSG	1689		
Db	1063	-----	NVNP	TKAOLADKVQ-----			1109		
Qy	1690	TALPSYRK	FVTKSRKSFIVLP	NDLKLARKG	IREVYFN	YNAKPALD	WYPSPRT	1749	
Db	1110	TAIDG-----	AHNNAE	VEAKNAE	IKIE--AIQ	PATTTDNMAKQAI-----	1149		
Qy	1750	PEITWRV	LQTVKSLAGV	LSMLRLW	MSLAWDMAAK	VPPGGSG	STRTETSE	IEITTEII	1809
Db	1150	-----	ATKANERK	TAIAOTODI				1166	

```

QY 1810 KRBDVPGYIGREYCIKICPI-----GVPEPKETPPQKGLR---- 1850
Db 1167 TAEETIAANANAVQANNNIENANSONDVQAKTGEASIDQVTPVKNKATAVTDA 1226
QY 1851 ----SSALPRKRPPEPKOGPVITIEFWAEFEELMEITAFARVEKEKAQAVEQAKKR 1906
Db 1227 KNNITPAATDNGVDPAKDKKNSIOS-----TOPATAVKSNAKND 1266
QY 1907 LEOQKPTVATSTSTSTSTSPKQVMAVPIGSVTGTKMVLTKVGSPTAVTFQ 1966
Db 1267 VDO-----AVTTQNDADINTGATTEK-----NAKDLVLKAK----- 1300
QY 1967 QKKNFHQFATVWKOGOSNGVVOQKVLGIIPSSTGTSQOTFTSFOPRTATVTIRPNT 2026
Db 1301 -EKAYOD-----ILNAQTNDVTQIK-----DQAVADVQGITADFTIKDVA 1340
QY 2027 SSGGTSNSQVITGPQIRPGAVITRPLQOSTLGKAITRTVMPVPGAPQOVMTQIRG 2086
Db 1341 KDELAKRAEQ-----KALIAQTADATTEKEEQANQOV--- 1373
QY 2087 QPVSTAVSAPNTVSTPGOKSLTSATSTSNI---OSSASQPPRPOOGVYL-TMAQLTOL 2142
Db 1374 -----DAELTQGNQNIENMOSIDVNTAKDNAIQALDPIQASTDVKTNARABELL 1422
QY 2143 TQGHGNGOGLTVVIOGQGTGOLIPQGVTVLPQPGOQLMQAMPNCTVORFLPTPLA 2202
Db 1423 TE---MOKNITEILINNNETNE---EKGNDI--GPVRAVEEGLN----- 1460
QY 2203 TTAITASTTTTSTTAAGTGEOROSKLSPOQOVHODKTLPPAOSSSVGPRAKQPTQAP 2262
Db 1461 ---INATTTTGV-TTAKDTAVOK-----VOOLHANPVKPKAGKTALDOAAADKKT-QI 1509
QY 2263 SARPOFOTOPOSAPPEVOTOPEVOTQTVSSHVPSAQPTHAOSKPOVAOSOPQSNV 2322
Db 1510 EOTPNASQOEINDAKQEVTELN-QAKTINDOSSTDEYVDNAVKEGKAKINA-VKTFSEY 1567
QY 2323 OGOSPVVOSPSQTRI-----RPSTPSQLSPQOSQV-----QTTTSOPIRPIQ 2365
Db 1568 KKDALAKIEAAVNAKVTADNSNASTSEIAEKOKLAEIKOTADQNVNQAATSKDDIEVQ 1627
QY 2366 PHTSL-QIPSOQOPQOPQVOSTQTLSSGQTLNVSVSSPSRQLOIQOPQOVIAVPO 2424
Db 1628 IHNDLNDINIDYITPTGKKESATTDLYAVADOKKNNISADT-NATODEKQO-----AIKO 1680
QY 2425 LOQOVQ-VLSQIOSQV---VAQIQAGQSGVPQIKLOLPI---QIOSSAVOTHO-I 2473
Db 1681 VQONVOTALESTINGVNDGDVDALTOGKAIDAIOVDATVVKANQVIDAKAEETKEST 1740
QY 2474 QNVVTVQAASVOEQLOQVQOLRDOO-----QKKQOQTEINVTSPSKLLIKV 2520
Db 1741 DOSDQLTAEKTEALAMIKQITDQAKGITDATTTAEVEKAKAQGLEAFDN-----IQI 1794
QY 2521 EIIQOVVKKHNAVIEHLKOKKSMTPAREENORMIVCNOVMKYIIDKID----- 2570
Db 1795 DSTEKO-----KAIEBELT-----ALDOIEAGVNVADADAT 1824
QY 2571 KEKQOA-----AKKRKEESVEOKRSKONAT-KLSALLFKHKEOLRAE---ILKKRA 2618
Db 1825 TEKEKFTALIEDISKATEDISDQTTNAELIAVKNAL---EOLKAORINPVVKKNA 1879
QY 2619 L-----LDKDQIEVQOEELKRDLLK-----KEKDLMLQATATAVAAPCPPV 2660
Db 1880 LEAIREVNVK--QIEIINKADADASAKELARTDGLRFPDFADKLDKQTQNTNEVALQNV 1937
QY 2661 TPVLPAPAPPPSPPPPGVQHTG 2684
Db 1938 T--IPAIEAIVPONDPNANDFTNSG 1959

```

THIS PAGE BLANK (USPTO)

1595 VSTSVSDSTASSTSESASTSTSVSDSNASTSTSESTSTSLDSTSKST-----1443
OY 1971 FHQFTWVWQOGSNGVVOQOKVLGILPISYGTSGOQRTTSFOPRTAT-----VTIR 2023
Db 1644 -SBSASTSTSESDBASSTSLDSTSTSVSESTSTSTSVSASNSSTSTSLDSTSKSTSL 1702
OY 2024 PNTSGS---GCTTNSQVITGPQIRPGMTVIRPLOOSTLGAIRTPVWVGAPQOVM 2080
Db 1703 DSTSTSTSEGSTSTSE-----SDSDASTSTSEST-----S 1734
OY 2081 TQIIRQPVSTANASANTVS---STPGQKSLTATSTST-----NIOSSAQPRPOGOV 2132
Db 1735 TSTSDSTSTSDASTSMVSDSNASTSTSLDSTSTSVSDSTASSTSEASTSTKSTSES 1794
OY 2133 KLTMAOLTOLOGHGNGOGLTVVIOGOGTTGOLQILPGCVTLPGGQLOMAAMPNGT 2192
Db 1795 TMASTSTSESTSVSDSTSTSTSDASTSTSESDNSESTSLSESTSTSVSDST-----1850
OY 2193 VQRFLETPLATTAATTAATTTTAVSTTAAGTGEORQSKLSPOMOVHODKTLPPAOSSSGVP 2252
Db 1851 -----ASTSASASTSTSVSDSNAST-----SLSGSTSTSVSD 1883
OY 2253 AKAQPTAPASARPOTQOPQSPAOPEVOTQPEVOTQTVSSHVPSEAOTHAQSKPQV 2312
Db 1884 STSTSTASASTSTSESDSDASTSLSGSTSTSTSDSTSTSDASTSTSEASTSTSV 1943
OY 2313 AAOQOPNOVQOGSPYVPOSPOT---RIRPSTPQSLSPGQOQOYOTTSTPPIQPH 2368
Db 1944 SEDSDSTSVSESSSTSVSDSTSTSTSEASTSTSESESTSESTSVSESSSTSTSDSSST 2003
OY 2369 SLOIPSGOPQOPQOVSSTQTLSSGOTLNOVSVSPSR 2407
Db 2004 STMSSTSEFTSQSPINSESOFT---GDSLSEDTITVQSK 2040

RESULT 3
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.9% Score 441.5 DB 4 Length 3696;
Best Local Similarity 17.1% Pred. No. 2.1e-17;
Matches 471; Conservative 4/0; Mismatches 1066; Indels 749; Gaps 106;

OY 311 QKNRPYIRHEPICVD-RSRKRYWFLNRR--LITEEDTENENEKKIWWYSTKVOLAELIDC 367
Db 569 QVAKDII---PSNYTLASYKYNKIKERAKQTVLDEETNNP-----606
OY 368 LKQDYEAELCKLLEKREIEIHRHMODITEDLNKARGSNKSLAANEELLESIRAKKD 427
Db 607 FNGRYSQTOIDDLHETQTLILNRVSASREINDKA-----QEMTDAY-----648
OY 428 IDWVKSPEETEXKNEFENDSKAEKNREFFEOQSLKOSDDKTPDDPQOGK-----480
Db 649 -----YSTELTTEEKDTLVQDLENHKNENISNNIDELTDGQVERKVEAGLHTL 697

OY 481 -SEEPTEV-----GDKNSVSANLGNNTINATSE-----ETSPSEGRSPVCL 522
Db 698 ESDTPHPVTPKPNAROVYNNRNDQOKTLIRNNHEATTEEOHEAIRQVHAHSDAIKIG-E 756
OY 523 SEFPDSSNMAEKVYASLLPDQPEEPKTKCESSNTSATTSTIOPNLENSMSSE-----576
Db 757 AETDTTVNEARDNGTKLIADV---PMPTRKABARAAVNTASNSKIKDINNNTQATLDER 813
OY 577 -----LNSOSE-----SAKAAD-----PENG-----ERESHTPVSIQREIV-----609
Db 814 NDALALNKRKDALQININATQGNNDVTEQNNGTNTIQVPLTPVKROMAIAFINAKAD 873
OY 610 -----GDFTEKSTGELSESPGAGSGASTRIITRLRNPDSKLSQLSQOYAAA 659
Db 874 EOKRLIOANNATTEEKADAEKRVYNEAV---ITANQITNATNTRVDQAOQTGGSGITSA 930
OY 660 AHEANKLFEKGG-----EVLVNSQGEISLSTKKEVIMKGINNNFKLGQEGKYRYHN 714
Db 931 ISPATKIKEDARAAVEKAKAIANOQINSNMATTEE---KEDALNOVEARHQAIAATINQ 987
OY 715 QYSTNSFALNK-----HOHREDHDKRHLAKHFCITPAGFEKMGNSVSGKVLITSTL 767
Db 988 AOSTQOYSEAKNNGINTINODQNAVKKNNT---KTILEQGNK-KSAIQOTPDATTEEK 1044
OY 768 RLTTIOLENNIPSSFPHPNASHRAMWIKAVQWCSKREFALALILECAVKPYMLPIW 827
Db 1045 QEAVSAVQAQVNTGITHINQANSND-----VDQELSNABQILTHTNVNVQKKPQA 1095
OY 828 REFLGHRRLHRMSIIEEKEKVKKKK-----KQEEELMQOATWVKYFPVKHGV 879
Db 1096 ROALIAKTNEKQASINDNGTILEEKQKAIQSLINDAANLADDEQITQAA-----1143
OY 880 WKQGEERYVTGYGGMWISKTHTYRFPVKLPQNTVNYRKSLEGTKNMNDENDESCKR 939
Db 1144 ---SNQVNDALNIGISINIKIOT-NFTKKQOARDVQNFQKEKELNTPHATQDEKQ 1199
OY 940 KC---SRPKRIKIEPDESEKDEYKSDAKGADONEMDITEKKQDQVYKELSDSDSK 996
Db 1200 DALTRLQAKETALNDINOQOTNONVDTALTSQIONQIONQVNVKRKQKEAKTTI-NDIYQ 1258
OY 997 PCKEEPHEVDDDKTBESHV---NQOESSQVNVN-----YSG-----FHL 1034
Db 1259 QHKOSIQONNDATTEKEVANNVMAQOVNSKIDNATTNNOIDIGVSGROSINATP 1318
OY 1035 RTSYK-----KTKSSKLDGLE-----RRIOFTLEEKORLEKILEGG 1074
Db 1319 DTSIKRRAKNDIDIKADKKIKTORINDATDELOEANKRIIEKAIKPAKINQNRSTROQ 1378
OY 1075 IKGIGKTSNSSKNLSSEPYI-TRAKEGCQSDSMRQ---EOSPANNNDQPEDLIQGCQ 1129
Db 1379 VNEAKTNGIKIENITPPTATVKSARQAVQONKANQEOINHQIONPDPATNEKQKAINVSA 1438
OY 1130 SDSSVLRMSDPSTTKNLYKDKRYLD--VSIRSPETCKQNSIENDIEKYS-----1181
Db 1439 ELARVOAQINAEHTQGV---KTIKDQATISLRINAQVVEKESARARAIQKATQOQFT 1495
OY 1182 -----DLASRQDEPTKSK-----TKGNDFPIDDKSLASADDIGTLI-----1217
Db 1496 NNNDNATDEKEVANNLVITKQSLDINISLSMND--VENAKVAGINIAVLPAITAY 1553
OY 1218 -CRNKPLIOESDPTTVSSSKALHSSVPKSTNDROAT-PLSRAMPFEGKIGCDSE--SN 1273
Db 1554 KSKAKKIDQKLAQOI-----NQIQHTQATTEKEKAQIALQANQKSEARATAIQNEHSNN 1608
OY 1274 STLBNSSDIYS-----IQDSSSEDMIVQNSNESISEQF-----TRQDYEVLEPLK 1320
Db 1609 GVAQAKNSNGIHEITELVMPDARKKSDAKQSIDNKRYNEQSNTINTTPPATDEKQKALDKLK 1668
OY 1321 CELVSG-----ESTGNCEDRLPYKGTGAN-CRKRPSQOKLEER-----1357
Db 1669 IAKDAGINKVDAQOTNOQVSDAKTEAITDTITNIOANVAKPSARVELDSKFEDLKROIQA 1728

OY	1338	-----	PVKKCDQJIKANTTDDKKNNEBSEKKG-QRTSPQI-----	1339
Db	1729	TPNATEEEKODAIORLGRKDEYKKNLINOBRUNEYQKHNGLOLEETIHANPTKSDA	1788	
OY	1395	-----	NGKDNKPRILYLGCLSEISESHSVSGNVEPKYN-----	1439
Db	1789	LOELQTKPISQTELLNNKNDATNDEKDEAKRLLEIKNKTITININQAQJNNVDANKNG	1848	
OY	1430	---INKTIIPENDIK---SLTYKESAIRPFING--DVIMED-----FNEINSSETKSH	1473	
Db	1849	MNEIATITIIIPATIKTDAKTAIDKRAEOQVYIINGNDNDADEKAEARKLVEKAKTEASNN	1908	
OY	1474	LLSSD---DAEGANRDSLELTPSRKESDSTQTTPTSPASCSPESNSVQVDEMEIETSEYKK	1530	
Db	1909	ITNSDTELEVNGATNLEKININIOBSTQTK-----MAQOINDKQOE--QLIO	1956	
OY	1531	VTSSPTISEESNLSDNFIDENGLPIKNENNVGESKRRVTITEVYTMSTIVATESKTVI	1590	
Db	1957	INNPDPATEEEKOEQATNRVNA-GI-AQALIONNHAJSTOEJNESKTSINSTATIKSVQPMVI	2014	
OY	1591	K-----VEKDKOTVVSSTEN-----CAKSTV-TTTTYTKLSPSTGSDVI	1633	
Db	2015	KKPRAINSLODEANNOXKTLIGNOGNATDDEKEAKOLVOKINEQOKIHESQTQNDQV-	2073	
OY	1634	ISVBEOSKTVVTTVTYS-----LTTGGTGLVTSMTVSEKSTRPKVLMKFSRPK	1684	
Db	2074	-NVKQAITAIKILINNAHAKRODAINILTVLAEKSKSDIRANQDATTSEK-----	2122	
OY	1685	KTRSGTALPSYRKRYVYTKSKSLFVLPBNDLKKLAKGIGREVPYFNAYAKALDIWYRP	1744	
Db	2123	---NTALQST-----DITLQAR---NNINGANTVA-----	2147	
OY	1745	SPRPFGITWYRBLQYFKSLAGVSLMLRLMLASIRMDMAKAVPBGSGSTRTESETEIT	1804	
Db	2148	-----LVDENLEDGOKQLOIRYLSYQTKQAKADIAQIIGQRBSTIDQNMNA	2194	
OY	1805	TT---EIKR-----RDVGPYGT---REFYCI	1825	
Db	2195	TTEEKQALELELNOETNGVNDRIQALANOMYDEKNMILETIRNVEPIYIKPRANEIT	2254	
OY	1826	RKIIICPIGVNPEPKETPTPOKGLRSSALPKRPETPKQTVGIJETWAEELIMEIR	1885	
Db	2255	RKKAEOQTLLINQODATLEKQDALGKLEVENKALNQ-----VSQHSNNVDKAIENN	2309	
OY	1886	APAEVEKEKAAVEQAKRRLBEOQPYIAISTSPSTSTISIPKQKVAWAPIS--G	1943	
Db	2310	GIATISEVHEPTEIIRKRAKOEIBQDOSQIDTINANKSTINEKSAIDRVNAKIDAIN	2369	
OY	1944	SVTTGTGMVLITTKVSPATVTFPOONKNFHQTPATVWKQOSNSGVVQOQKVLGIIPSS-	2002	
Db	2370	NITNAT---TTOLVNDA-----KNSGNTSISO---TLPSTA	2399	
OY	2003	--TGTSOQTFSPQRTAVTIRPNTSGSGGTTNSQVITGPQIRPMGTVIRTPLOQ--	2057	
Db	2400	VKTNALALALASEAKNKMAIIDQTPNATAEKEEANKV-----DRLOEAD	2445	
OY	2058	STLKALIIIRPY-MYOGAQPO---YMTQILBQRPSTVASAPNVVSSTPQKSLTKSTS	2113	
Db	2446	ANILKATTEDEVNNTKQAOYONINAVOEVIKQNKQNLN-----QPLDNQKTIENTP	2500	
OY	2114	TSNIQSSASQPPRPOQOVYLTMAOULTOLLOGHGNOGLTVYIOGQGOTTGLOLILPGV	2173	
Db	2501	DATLEEKA-EANRLLOVNLVSTSDELIANVDHNEVNDALD-----KARPIE	2546	
OY	2174	TVLDFGQOQLOQAMRPGIVQRLFTPLATLTATSTTTVYSTTAAGTGEOROSKLSPO	2233	
Db	2547	EIVP---QVSKKRQVLNVAIDQAF-----NSQOIEQOENQOATNEKTEALKKI	2591	
OY	2234	MC-VHOOK-PLTPROOS-SVGPKAQOPQTAQSPARPOQTOPQSPAQEVQTOPQEVQOT	2290	
Db	2552	NOLLQAKVINDAQSKVDVSAKTSIODIEDIOHPHOTKATGHRILNEKAN--QOOS	2648	
OY	2291	TVSSHVSEAOPTHAQSSKPO-----VAAQSQPOSN-----YQOGSPVAVQSPQIR	2347	

Db	2649	TTAHPNSTIEEROASAKLOEVLKKAIAKIDGOTNDDEKTVYNGIAEINILPA	TT	2707
Qy	2338	IRPSTPOLSFGOSOVOTTTISOPILPHTSLPISQSGOPSOPOVOSTOTLSGGTL		2397
Db	2708	VKDKAKADVNAKEOK	-----NLQINSNDTEATTEKVLASDNLNHVETT	2752
Qy	2398	NOVSSSSPRQLOIOOP	-----QPOYLAIPQLQOQOYVLSLOISOQYVAQLOIAQOS	2448
Db	2753	NOALJEDAVTINQVNVENKKGITGRIDPILPVAKKPAK	-----SKIESAV	2797
Qy	2449	GVPOOIKLOLPIDIOGSSAAVOTQIOINVYTVQAASVOEOLQFOOL	-----RDQOORKKQ	2503
Db	2798	-----EKKTEINOQNTATHDE	-----VREGLOLQNIHEKAKNDVNOQSTN	2839
Qy	2504	QOIEINWTPSKLLIKVEITQOQVAKHNAV	-----TEHKOKKSKSTPAERE	2550
Db	2840	QOQVE--NAEONSIDQINRPFDSKRRNAVEIYKAQONKIDIEOFESSTOEBEKDAL		2896
Qy	2551	-----ENORMIYCNQVMKY	-----TLDKIDKEEKQAA	2577
Db	2897	QHLDQVKEIINSINQANTDNEVDNAKTSGLNITTEPREYKKNKNAILLKIDVSDPQEA		2956
Qy	2578	KKRKRESDQOKSKONATLSALLFKHKEOLPAELIKKRALLDKLOIEVQEBLK		2633
Db	2957	IINGPATDELOEAN-SLKINKILLDARKQOIGLAHTHNEV	-----DDIYENQSQMRK	3008

```

: RESULT 4
: US-09-134-001C-3159
: Sequence 3159, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3159
: LENGTH: 10182
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3159

```

[illegible]


```

Db 4909 -----LIPN-----LFWQKAKEKELVNSQTRREV 4933
QY 2233 QMVOHQDKTLPPAOSSSVGPAAK---AOPQTAQPSARPOQTOPOSPAPQEVOTQEVOTQ 2289
Db 4934 QEOINQAASL---DSSGKGLSLVAKOPTVQKTSYINEDQPSOAVNDSIT---MGQ 4985
QY 2290 TTVSSHPVEAOPTHAOSKSPQVAAQSOQPSNVQGSFVQSPQTRIRSTPSOLSFG 2349
Db 4986 TIINKTADPVLDKTLVNDNAISINISTKENALHGEQKLTAKTEALNAL---NTLADLNTP 5041
QY 2350 QOSOVQTTTQOPPIQPIQTSIQIPQSQPOQOP---QVOSTQOTLSSGOTLQNVSVSPSRP 2408
Db 5042 QKEAIKTAIIN---TAHTRTDYTAE---QSKANDINSAMHTLQONISDNE-SVTNESN- 5091
QY 2409 QLOIQOQPOQVIAVPOIQOQOVLSQISQVYVQIQAOQSSVPOQIKQLPIQIQOQSSAV 2468
Db 5092 -----YNAEPEKOHAFTEALNNAKKEIVNE-----QOATLDANSINOKQAOI 5133
QY 2469 QTHOIQNVTVQVQASVOEQLQVQOLRQOQKQOQOIEINVTNPSKLL---IKVEIIQ 2524
Db 5134 LT-----TKNALDGEQLRAKENADQELINTLQULDAQRNBEKGLINSQTRTEVAS 5186
QY 2525 KOVVMKH-NAVIEHLKQ---KKSMTPAER---EENQRMIVCNQVMKYIIDKTDKEEQ 2575
Db 5187 QLAQAKELNKVMEQILNHLINGKNOMINSKFTINEDANQOQAVNAISAEALK-NKSONP 5245
QY 2576 AAKRRKRESEVEQKRSKONATKLSALLFKHKEQLRAELIKRALLDDLOIE---VOEEL 2632
Db 5246 ELQVYTEQALINNINSAINNINNGEAKLTKAKEDAVASINNLSGITNEQIKENOAVNQAO 5305
QY 2633 KRDLKIKKENDLMQAOA 2650
Db 5306 TRDQVANKLRDAEALDQS 5323

RESULT 5
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djahail, Malek
; APPLICANT: Selleri, Licia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

Query Match
Best Local Similarity 17.6%; Pred. No. 7e-13;
Matches 620; Conservative 405; Mismatches 1211; Indels 1278; Gaps 153;

QY 5 EEEEDDDAEETQDSEDEDEDEDDDDSDYPEMEDDDDDASYCTESS-----54
Db 855 EELSKRDADQSVKDKSRERDREREKENK--RESRREKRKKGSEIQSSSALYPVGRVSK 912
QY 55 -----FRSHSTYSSTPGRKPRVH-----73
Db 913 EKVVGEDVATSSSAKAKTGRKSSSDSGDITSVLTGDTTAVKTLIKKGRONLEKTN 972
QY 74 ---RPSPILEEKDIPLEPPKSSSEDLMPVNEHIMNVIATIEVLRNFGVLRSPREFED 130
Db 973 LDLGPTAPSLKEKTLTLPSSS-----TYKHSSTISGSHL-----1009
QY 131 FCAALVSQEOCTTMAEMHYVL-----KAVLEEDTSNTTF- 166
Db 1010 -----AQADKLPTDTRVAVSLKKAQAQCKTEKSKLQOTDQPKAQGOQSDSSTSVR 1063
QY 167 GP-----ADKQSVNSTLYFLDGMTPPE---VLRVYCESDKREYHHVLP 206
Db 1064 GPRKHVCRRRAVALGRRKRVFPDDMT---LSALPWEREKILSSMGNDSS--SIAG 1117
QY 207 YQEAEDYPYGPVENEKIKVLOFLVDQFLTNTIAREELSEGVIOYDHCRCVHKLGDLCC 266
Db 1118 SEDAE--PLAPPIKPIK-----PYTRKAKAQEPVKKGRSRRCGQ-----C 1157
QY 267 ETCSAVYHLECVKPRLEVPDEMOCEVCVAHKVPGVTCVALETQKKPYIRHEPIGYDR 326
Db 1158 PGC-----QVPEDE---CGVC-----TNCI-----DKP-----1176
QY 327 SRRKWFNLNRLIIEEDTENENEKKIYVSTQVLAELIDCLDKDVEALCKILEEMRE 386
Db 1177 ---KFGGRN-----IKQCKCKMKCONLQWMP 1201
QY 387 EIHHRMDITEDLTNKARGSKSFLAANBEILSIRAKGDIQNVNPSPEETEKDKETEN 446
Db 1202 KAY-----LQQAQAVKK-----KEKSKTSBKSKDSESVVKNVDS 1239
QY 447 DSKDAENKREFEQQSLKSDDKTPDDDEQCKSEPEVEGDKGNSVSNLGDNTNAT 506
Db 1240 SQKTPSARDE---PAKKSSSEPPPKPVEEKSE---GN-VSAP-GPESQAT 1286
QY 507 --SEETSPSEGRSPVGLSETPPDSNMAEKKVASELPQDVPBEPNKTCESSNATSATTST 564
Db 1287 TPASRKSSKVQSPALYIIPQPTTGPBK-----EVKKTTPSEPKK-----1329
QY 565 QPNLENSNSSSELNQSQESAKAADP-----ENGRESHTPVSIQELIVGFTSEKST 618
Db 1330 QP-----PPESGPEQSKOKKVAAPRPSIPVKQPKPEKEKPPVKNQOEN-----1372
QY 619 GELSESPGAKGASGRITRLRNPDPSKLSQQAQVAAAHAANKLFPREGKEVLVNS 678
Db 1373 -----AGTLNLTSLTSLNGNS---KQKIPADGVHRIYDFEDECDAENWVE 1415
QY 679 QGEISRLS---TKKEVIMGNINNYFKLQGEQKYRVYHNYQSTNSPALNKHQRDHDH 734
Db 1416 MGLGLITLSPITPRVVC-----FLACSSHVEVEYQVCCERP-----1454
QY 735 RRLHAKFCLTPAGEFKMGNSVGSKVLITSLRLTITQLENNIPSPFHPMASHRANW 794
Db 1455 -----HKFCL-----ENRPLEDDQEMNCRCRC-- 1478
QY 795 IKAQVQSKRERLALALILEC-----AKPVVLMPLMREPLIGHTRLAR 838

```

Db 1479 -KCHVCG--RQH0ATK0LLECNKCN5HYRCELGPNYPRKPKKKKWM1-----CTKCVR 1531
QY 839 M5TIEREEKYKKKKKEKOEETMO0ATWVVKYTFVKHQVMK--OKGE-----885
Db 1532 C5KSCSGTTPGK-----GMDAQW-SHDFSLCHDCAKLFKAGNFCPLCDKYDD 1577
QY 886 ---EYAVTVGGMS-WIS-----KTHVYRFVVKLPGN---TWVNRK-----920
Db 1578 DDE5YMMOCCKGCDRWVH5KCEMLSD5EMYEIL5NLP5VAAYTCVNOTERHPAEMRLALEK 1637
QY 921 ----SLEGTNNMDEMD5DKK5R5PKKIKIEPDS-----KDEVGSD-----AA 965
Db 1638 EL0ISLK0VLTALN5R5T5HLIR0AKPRPLN5P5TES5IS5R5SP5CP5DP5VLE5S 1697
QY 966 KGADQNM5DI5KITEKKDD-----VKEL5SD5DK5PCKE5EPM5VDDMK 1010
Db 1698 KODDQ0PLDL5GVK5RKM5DGN5T5VLE5FSDI5KI0A5IN5DGGP-----EIKK 1748
QY 1011 T5SHVAC0SSQV0VYVNV5EGFHLK5S-YKKK5SK5LGL5ERIK0FTL-----1060
Db 1749 AN5M5V5SFT5R0M5--RVP5P5SVK5K5R5W5EP5K5V5SNG5M5LP5NAL5P5LDHN5YAQ05R 1806
QY 1061 -BEK0RLEKIKLEGGI-----KGIGK5T5N5SK5N5S5P5ITK5K5EG05DSM5R05Q5P 1113
Db 1807 RE5N5HT50P5PLMKK5IIP5PK5K5G5P5D5SP5T5PL5HP5TP5IL-----STD5R5K5D5F 1858
QY 1114 NAN-----ND0P5DLI0G-----CS05DS5V5LRMSD-----1139
Db 1859 ELN5P5PGI5DN05CAL5T5YGD5DSAND5AGRL5YIG0N5M5THV5CAL5MS5AEV5EDD5G5SLK 1918
QY 1140 -----PSHTTNK5LYPRD5VLDDV5INS5P5ETK5CKR0NS 1171
Db 1919 N5HMAVIRK0L5CE5C0K5PGAT5VGC5L5T5C5T5N5H5F5M5SR5AK5N5CV5FLD5KK5YV05COR5HRD 1978
QY 1172 I-----ENDIE-----EKV5DLAS--RG0EPT5K5TK5GN5F5IDD-----SKLAS 1209
Db 1979 LIG5EV5P5ENG5E5VR5K5RV5VDE5FG5ISLRK5FL5GL5EP5N5H5M5IG5MT5D5C5IL5DL5SD 2038
QY 1210 ADD-----IG-----TLICK--NKK5PLI0-----E5ED5T5V5S5 1236
Db 2039 CEDK5LP5IG5C5R5V5W5T5TD5AK5RCV5Y0K5I5E5C5RP5V5EP5D5INST5VE5D5ERT5IAH5P 2098
QY 1237 K5ALH5SV5K5STN-----DR-----DAT5PL5R5M5D5F--BEK5L5G 1268
Db 2099 T5P5T5ESS5K55E05T5A5E5I5PP5P5D5R5P5H5S05SG5CY5YH5ISK5V5P5IR5T5P5S5P5T05R5P5G 2158
QY 1269 D--SE5N5T5L5E5N5SD5V5S5I05D5SE5DM5I05N5NE5I5S05F5R5T5R05Q5V5LE5P--LK5CELY 1324
Db 2159 RPL5P5AG5PT5PT5T5H5E5I5Y5GDP-----LL5SG5LR5IG--SRH5T5SL5P05R5KL5R5M 2209
QY 1325 SGE5T5G5N5CD5RL5PV5K5GT5E5ANG5KK5P5S0K--K5LE5RP5N5K5C5D05IK5L5N5T5D5KK5N-----1377
Db 2210 SP5R5T5G5N5T5SR5NN5V5SV5T5GT5AT5D5ESS5AK5YV5D5H5L5G5PL5N5T5SL5G05MT5S5N5L05R5Y 2269
QY 1378 ---E5R5E5SK5K05R5T5F-Q5INGK5DN5K5P--I5YK5G5CL5K5E5SR5V5SG5N5VE5P5K5V5N---1429
Db 2270 V5YV5G5N5N5H5L5D5SS5SEM5K05S5AD5V5SK5S5SL5K5G5E5K5T5V5L5-----SK5S5E5S5AH5N5V5AY 2324
QY 1430 --INK5I5P5E--NDIK5L5T5V5K5ESA5IR5P5IND5V5I5M5ED5N5ER5N5--SE5K5H5L5SS5D5A5G5N5R 1485
Db 2325 P5G5I5P5K5AP5V5H5T5S5K5EL5V5K5I5G5FA5EP5SV--SF5SK5E5AL5SF5P5HL5R5G05R5ND05HT 2382
QY 1486 D5LET5L5P5K5ESD5--T05T5P5S5A5C5P5E5N5V5N0-----VED5E5I5E5T5E 1527
Db 2383 D5T05AN5SP5DD5T5E5K5T5KL5G5M5NR5S5I5NE5H5M5SS5R5DR05K5CK5K5E5F5FK5H5S 2442
QY 1528 V5K5V5I5S5P5T5E5E5N5L5AD5P5D5E-----NGL5P5INK5E 1560
Db 2443 K5F5LE5G5V5T5G5E5G5IK5E5F5E5V5L5P5E5M5G5R5C5NN5V5SD5K5IG5D5K5GL5M5G5V5P5AP5M 2502
QY 1561 N5NG5E5K-----R5K5V5I5E5V5T5M5T5V5A5E5K5T5V5IK5E5K5Q5V5S5T5E5N5A5K5T5V5T 1614
Db 2503 QV5E5G5AK5E5L05AP5R5K5T5V5T5L5P5L5K5M5E5S05N5AL5K-----ESS5P5ASP5L05E5T5S 2553

QY 1615 T5T5T5V5K5L5T5P5T5G5G5VDI5I5V5K5E5K5V5V5T5T5V5D5L5T5T5G5T5V5T5M5T5V5K5E5Y5TRDK 1674
Db 2554 P5T5E5I5A5E5N5P5D5G-----P5VAQ5P5PN--N5T5C05Q5S05N5N5Y0NL-----FVQDR 2595
QY 1675 VK5M5K5R5K5K5T5S5G5T5AL5S5Y5K5F5V5K5T5K5I5F5V5P5ND5DK5L--ARK5G5I5R5E5V5YF--1730
Db 2596 -N5L5P5D5P5K5P05D5G5F5K5R5Y5P5R5S5AR5ANS5NM5F-----GL5P5LY5G5R5Y5G5E5D5I5P5Y5S 2649
QY 1731 -----N5NA5K5PAL5D5I5P5Y5SP5P5T5G5I5T5W5Y5R5L05T5V5K5SL5AG5V5L5M 1771
Db 2650 ST5G5K5R5K5K5S5A5E5Q5V5G5AD5L5T5S5ED5DL5Y5Y5N5F7-----TV5SS5G5E-----2694
QY 1772 R5L5M5A5L5R5D5M5A5K5V5P-----GG5ST5R5E5T5E5T5E5T5I5I5K5R5D5V5P5G5I5R5E5Y5C5IRK 1827
Db 2695 R5L5A5N5L5F5R5E5E5C5D5L5P5K5I5Q5LD5G5VD5G5T5-SDT5V5T5A5T-----2732
QY 1828 I5C5P5I5G5P5E5T5K5E5T5P5P5R05K5GL5R5SA--L5P5K5R5E5P5K05T5G5P5I5E5T5W5A5E5E5L5E5M5E5R 1885
Db 2733 -----TRK5S05I5P5R5K5G5E5N5E5L5K5ID5R5D5A5G5E5K5E5H5T5K5S5V5G5H5K5N5E5K5M5D5N5H 2783
QY 1886 A5P5A5R5V5E5K5E5K5A5Q5A5E5Q5A5K5R5L5E5Q5K5P5V5I5A5T5S5T5P5T5S5T5I5S5P5A5K5V5M5A5P5I5G5AV 1945
Db 2784 SV5R-----VK5T05G05D5L5E5A5Q5L5SL-----ESS5R5V5HT5T5P5DK5N5LD-----2821
QY 1946 T5G5T5K5N5V5L5T5K5V5S-----PAT5V5T5Q5N5K5N5F5H05T5P5AT5W5V5K05G5NS-----G 1987
Db 2822 T5Y5N5T5EL5K5D5SD5N5N5D5D5C5G5N5IL5P5D5IM5D5V5L5K5N5P5M5Q5AL5G5P5ESS5SELL5N5L5G5G5G 2881
QY 1988 V5V05Q5K5V5G5I5P5S5T5G5T5S05Q5T5S5F05P5R5AT5YT5I5P5N5T5G5S5G5I5T5S5N5Q5V5I5T5G5P5I5R5G 2047
Db 2882 L5D5N5R5E5K5D5M5GL5F-----E5V5S05L5P5T5E5P5V--D5SV5S5I5S5A5E05F5E5L5P5L5P5SD 2929
QY 2048 M5V5I5R5T5P5L05G5T5L5G5K5A5I5R5P5V5W5G5A5P5Q5V5W5T05I5R05P5V5T5A5S5A5N5T5V5S5T5P5G5K5S 2107
Db 2930 L5V5L5T5-----R5SP5V-----P5Q5N5P5R5L-----A5V5DS5G5E5R 2958
QY 2108 L5T5A5T5S5N5I05S5A5P5-----P5RP5Q-----O5G5V--K5L5M5A05L5T05L05G5H 2147
Db 2959 V5T--I5E5K5V5A5S5E5SP5ALL5SP5G5VD5P5T5P5G5H5M5P5H5F5I5G5H5M5D5A5H5I5S5P5G5SV5Q5G5H 3016
QY 2148 G5N05GL5V5I05G05G05T5G05L05L5P05G5V5-----L5P5G5Q-----2181
Db 3017 N5M05DL5F-----R5N5S5T5P5G5L0--V5V5P5T5V5I05N05K5V5N5D5S5P5G5P5O5I5N5A5V05T5P5H5L 3070
QY 2182 -----O5M5Q-----A5AMP5N5G5V05R5F5L5T5P5L5A5T5A5T5S5T5T5T5V5T5T5A5G5T5B5Q 2225
Db 3071 K5P5A5T5E5K5I5V5N05M05P5L5Y5L05L5P5N5G5V5T05K5I05L5T5S5V5S5T5P5M5E5N5T5V5L5G5P5G5-GLT 3129
QY 2226 R05K5L5P5Q5V5H05D5K5L5P5P5A05S5SV5P5AK-----A05P05A05P5AR5P05P05P-----2272
Db 3130 L5T5G5L5N5P5L5T5G5--S5L5P5A5S5K5G5L5P5M5H05H5L5F5P5A5T05S5P5P5N5I5N5P5G5L5I5G 3187
QY 2273 Q5P5A5P5A5E5V--O5T05P5V05T5I5Y5S5H5V5S--E5A5P--T5H5Q55K5P5V5A5A05P05Q5V5G0 2325
Db 3188 Q5P5P5D5Q5L5V5E5S5O5R5D5L5T5V5A5T--P5S5G5L5K5R5I5S5L05T5K5K5K5L5A5P5S5T5P5N5I--3242
QY 2326 SP5V5R5V05S-----P5Q5R5I5R5P5S5P5O5L5P5G-----Q05Q5V05T5T5S05P5I5P 2363
Db 3243 A5P5D5V5S5N5M5T5L5N5F5T5S0L--P5N5P5SL5D5L5G5L5N5T5S5R5Y5P5N5I5K5K5S5I5M5T5E5P5AP 3299
QY 2364 I05P5H-----T5L05P5S05Q5P05Q5P05V05S5T05L5S5G05L5N05V5S5P5R5P05L05I05Q5P-2415
Db 3300 L5P05V5G5I5A5A5A5G5T5I5S05P5H5L5T5S5V5G5L5A5S5S5V5L5N5V5M5Q5T5T5T5P5T5S5A5V5G 3359
QY 2416 -----O5P5V5I5A5P05L05Q5V5V5L5S05Q5V5V5A05I05A05Q5S5P5O5I5K5L05P5I05I05Q5SA--2467
Db 3360 H5T5L5N5P5R5L5G5T5P5D5I5G5-----I5N5L5L5I5K5A5Q5S5I5G5I05D0--F5V5AL5P5S5G5M5P 3406
QY 2468 -V05T05I05N5V5V5V5A5S5V5E05L05R5V05L5D05Q5K5K5Q5Q5I5E5I5N5V5M5P5S5L5L5K5V5E5I05K0 2526
Db 3407 Q5G5T5O5T5P5T5A5T5A5S5I5C5V5L-----P5T05T5G5I5T5A5S5P5-----3442

```

QY 2527 VVKHNAVIEHLKOKKSMTPAREENOMATYCNQVMKYIILDKIDKEKQAAKKRREESV 2586
DB 3443 -----GEAENHQLQAHVOLL----- 3458
QY 2587 EOKSKONATLSALLRHKHEQLRAELIKKRALIDLOLEVOELKRDLIKKEKDLMO 2646
DB 3459 -----ASKTGT-----HSSQ-----RDLDSAGPOVENFTQTVDPAPNSMG 3493
QY 2647 LAQATAVAAPCPVTPVLPAAPPAPPPPPPGVQHGLST---PTLPVASOKKREEEK 2704
DB 3494 LEQKKALS-----SAVQASTSPGSGSPSSQSRASASVPEPTPKPKTKRFLPLD 3547
QY 2705 DSSSKSKKKMIKISTSKET---KKDKLYCICKPYDESKFYIGCDRCQNMVHGRCVGI 2760
DB 3548 KGNCKKHVSHLRTSPSSSEAHIPDEETSLTSGTSGAEAE----- 3588
QY 2761 LOSAELIDEVCPQOCSTEDAMTVL-----TPLEKDEGLKRYLRLSLQAHKAMPF 2813
DB 3589 -QODTASVEOSSQECQOPAGQVAVLPEVQVTONPANEO----- 3626
QY 2814 LEVDPDAPRYGVVKEPMDLAMEEPVORRYEKLFEVADMTKIFDNCRYNPSDSP 2873
DB 3627 -ESAEPTVEEESNFSSPLMLQOQOKR---ESTTEKKPKKGLVFE-----ISSDG 3677
QY 2874 FYQCAEVLESEFYOKLKGFKASRSHNKLQSTAS 2907
DB 3678 FOICAESIED-----AMKSLTQKQVQARS 3701

```

RESULT 6

```

US-07-741-940-7
Sequence 7, Application US/07741940
Patent No. 5352775

```

GENERAL INFORMATION:

```

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARK
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SRO ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids

```

```

; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-07-741-940-7

```

```

Query Match 2.4%; Score 362.5; DB 1; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8.1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

```

```

QY 373 WEAECLKILEMREIREHRHDITFEDLTNRKAGSKNSKSPLEAANEELIESIYAKKGDIDNVK 432
DB 698 WDMGAVMLKLNHLSHKMT-----AMGS-----AALRLMANRPAKKKDA-NIM 742
QY 433 SPEET-----EKDKN-ETENDSKDAEKNEEPEQDSLEKSDDKTPDDDPQOKSEPT 485
DB 743 SPSSSLPSLHVRRKKALEAELDAQHLETFDNDINLSPKASHRSKQKHQSLVGDVYFDI 802
QY 486 EYCDKGNVSANLGD-----NTF-----NATSEETSPSGRSPVGC--- 521
DB 803 NRHDDNRSNDNFNTGNMTVLSPLYLNTTVLPSSSSRSGSLDSRSERKSLREREGIGLGNV 862
QY 522 --LSETPDSSNMMAKKVASLQ--DYVEEPN--KTCESSNTSATTTSI-----QPTLEN 570
DB 863 HPATENGTSSKRGLOJSTRAQIAKYMEVSAHTSQEDRSSGSTTELYCVDERNALR 922
QY 571 SNSSSELNSSQSESAAKADPENGERESHPPVSTOIEIVGDTSEKSTGELSESPGAGK 630
DB 923 RSSAAHTHSNTYNTTKS-----ENSNFTCSMPYA---KLEYKRSSANDSLNVSSDYGYKR 975
QY 631 ASGSTRLITLHRNDSKLSQLKS--QQVAAAHEANKL-EKEGEGVLVN----- 677
DB 976 GOMKPSIESTSEDESKFCSTGYQYRADLAHKHSANMDNDGELDTPIYSLIKSYDEQL 1035
QY 678 SOGEISRLSTKKEVIMKGINNVEFKLQEGSKYRYHNOYSTNSELNKHQREDHDK-RR 736
DB 1036 NSGRQSPSQNBRMARPHIIEDELKQSEQOSR---NQSTT--YPVYTESTDDKHLKFOR 1090
QY 737 HLAHFCFLTPAGEFKKNGSVHGSKVLTITSLRLITTOLENNITSSFFHPWASHRANWIK 796
DB 1091 HFGQECVSPYRSRANGS-----ETNRVGS-----HNGINQNV 1124
QY 797 AYQMCSPREFALALALIECAVRYVWLPIRBEFLGHTLHRMTSTREKEKVKKKEK 866
DB 1125 SQSLQOED-----DYEODKPTNYSERY 1146
QY 857 QEEETMQOATWVXYTPPVKHQVWKQGEERYRVYGGWSMISKTHVYREVPRLPGNTNV 916
DB 1147 SEER-----QHDEERPTNY-----SI 1163
QY 917 NYRKSLEGTNNNDENDES-----DKKCSRSPPKIKITIEDSEKDEV 959
DB 1164 KYNEE---KRHYDQPIDYSLKATDIPSSQOKSFSPKSSQSSKTEHMSSESTST 1219
QY 960 KGSAAKGAONEMDIDKITEKKQDVKELDSDS-----DKP-CKEPEMEV 1005
DB 1220 PSSNAKR---ONLHPSSAOSRSGQPOKAATKYVSSINQETIOTYCVEDPIFCSRSSL 1276
QY 1006 DDMKTESHYNCOSSQ-----VDVVNVSEGFHLRT-----SYKKRTSSKL 1047
DB 1277 SLSASADEIGCNGTTOEASANTLQIALEIKKIGTASADPYSEVPAVYQHRITSSRL 1336
QY 1048 DGLIERIKOFTLEKORLEKIKLEGIKIGTSTNSSN-----LSESPVI-YKAKES 1101
DB 1337 QG-----SLSSESARHKAVERSSGAKSPSKSGAOTPKSPENHYOETPLMFISCTSV 1389
QY 1102 QQSDSMQEQSPANNNOQPEDLIQCGSQSDSVILRMSDPHTTNKLYPKRVLDVDSIRS 1161
DB 1390 SSLSDFESRSIASSVOSEP-----CSGAVSGIISPSDLPSDQGWMPSPR----- 1434

```

QY 1162 PETKCPKONSTIENDIEKVSOLASRGQPTKSKTGNDFFIDSKLASAD----- 1211
1435 SKTPEPPPTQOTKRE-----VFKNKAPFAEKRESG-----PKQAAVNAAVOGVOLP 1482
QY 1212 DIGLI-----CKNK-----KPLIOESDITV-----SSKSAHSSVYK 1246
1483 DADTLHFATESTPDGFCSSSLSALSLDEPFICKDVELRIMPVEONDNONENESSEPK 1542
QY 1247 STNDRAATPLSRAMDFGKLGCDSESNSTLENSDITVSIOSSPEMDIVONSNSISBQF 1306
1543 ESNMNO-----EKEAKTIDSEK-----LDDSDDD----- 1569
QY 1307 RTREDOVEVLEPLKCEVSGESTGNCEDRLPVKCTEANGKKPQOQKLEPRVVKCSDOI 1366
1570 -----DIEILLE--EC-IISA-----MPTKSSR--KAKKPAQTASKLPPVARKPSQL 1611
QY 1367 ---KTKMTTKKNNESEKKGORTSTFQINGDNKPKITLKGECLKSESKVYSGNV 1423
1612 PVYKLLPSQNLQPOKHVSFTPG-----DDMPRYV-----EGTPPINFST 1652
QY 1424 EPKYNINIKIIPENDIKSLFVKESAIRPFGVIMEDFNRNSETKSHLLSSDAEGN 1483
1653 ATSLSDLTIESPNEL-----AAGEVYAGGAGSGEFEKRTIIPTEGR--STDEAOGG 1702
QY 1484 YRDSLETLPTSKES-----DSTOTTTTPSACPESS 1512
1703 KTSYV--TIPELDNKAEBGDLACINSAMPKKGSHKPFYVKKIMDQVOQASASSAPNK 1761
QY 1513 NSVQVDEDMETSEVKKYVS-----SPTSEESNLSDNF 1548
1762 NQLDGKR--KKPSVPVPIQONTERTYRVRKNKADSKNNLNAERVEYSDNKOSKQNLKNNS 1819
QY 1549 IDENGLPINKENENYGE-----SKRTYV 1571
1820 KDEFDKLPNNEDRYRGSAFADPSPHHTPIECTPYCFSRNDSLSLDDDDDDVDLSREK-- 1877
QY 1572 ITEVTTMTSTVAESKTVIVKEGDKOTVVSSTENCAKSTVT-----TTTTVTYK 1621
1878 -AELRAKAKENKSEAKYVTSHELTSNQOSANKTQAIKQPIRNGOKPILQKOSTFPOSS 1936
QY 1622 LSTPSTGSDIISVKROSKVVTYTTDLSLTGGLVMSYMSKEXSTDKYKLMKFS 1661
1937 KOLPDRGAAD--EKQONRAIENTPVCFSS--HNSLSLSLSDIOENNNKENEPIKETE 1990
QY 1682 RP-----KTRSGTALPSYRK-----EYTKSTKKSIFVLPMDD-----K 1716
1991 PPOSGQEPKRPQASGVAPKSHVEDTPVCFSRNDSLSLSLSDSDLLQECISSAMKXK 2050
QY 1717 KLRKGGIREVYPTFNNAKP-----ALDIMPYSPFPPTFGIT-----WRYRL 1758
2051 KPSRLKGDNE-----KHSRNNMGILGEDLTLDLKIDQRPDSEHGLSPDENFDMKAIO 2104
QY 1759 QTVKSLAGVSLMLRLMASLRMDMAKVPVGGSTRTETSETTEITTEIKRQDVCPYG 1818
2105 EGANSIV-----SSLHQAAAACL-----SKQASSDSD-----SILSLKS---G 2140
QY 1819 IREYCIKRIKIICPIGPVETKREPTPQKGLRSALRPKREPTEPKQGPVLIETWAESE 1878
2141 I-----SLGSPFHLPPDOEKFTSNKGR--ILKPGEKST-----LETKKIESE 2133
QY 1879 -----LELWEIRAFABERVEKEKAQAVEQAKRLEQOKP-----TVIATST 1919
2184 SKGIGKGYKVKSLITGKVRNSB-----ISGOMKOPLOANMPSISNGRPMIHIGP 2234
QY 1920 TSPSTSTSTIISPAQKVMAPISGVYTTGTRKMLVTTKVGSAIATYTPQONKNFHOTFATW 1979
2235 VRNSSSTSPVSKKGPPLKTPASSPSEGGTATTSPRGARPSV----- 2277
QY 1980 KOGOSNGVVOVOQKVLGIIPSSSTGSOQTFTSFOPRTATVTIPNPTSGSGGTTSSNOVI 2039
2278 ---KSELSPVARQTSOIG-----GSSK-----APSRSGSHDSTIPS----- 2309

QY 2040 TGPQIRPGMTVIRPPLQOSTLGKAIIRTPVWVOPG---APQOVMTQIIRGQPVSTAVSA 2095
2310 -----RPAQOGLSPBIPGSRNS-----ISPGRNKISPPNKLISQLPRTSSPSTA--- 2353
QY 2096 PNIVSTPQKSLTS-----ATSTSNIOSSASQPPRQOQOVKLTMAQLTQLIOGH 2146
2354 -STKSSGSGKMSYNSPGROMSQNLTKOTGLSKNASSIPRESAS-----KGLQNMNNGN 2407
QY 2147 GNGOGLVYVIGOGQOTYQGLQ-----LIPQGVTVLPQGOLOMAMPNGTVQRFLETF 2199
2408 GANKVELSRNSTRKSSGSESDRSERYLVNQSFIEKAPSPTRKLEESAS-----FE 2462
QY 2200 PLATTATTAATTTTTVSTTAAGTEOROSKLSP-----QMOVOD-----KTLPPAQSS 2248
2463 SLSPSSRPASTPRAQOTPV-----LSPSLPDKSLSTHSSVOAGCWRKLPNLS 2512
QY 2249 SVGPAKQOPQOASARPQOTOPOSPAPQPEVQOPEVQOTYVSSHVPSAOQTHAOS 2308
2513 TIEYNDGRPAKRHDIAHSHSPSLRPLINRGTWKRE--HSHKSSSLPRVSTWRTGSS 2569
QY 2309 KPOVAAOSQPSNVQOGSPVYVOSPSOTRIRPSTPSOLS--PGQOSOVQTTTSCPIPIQPH 2367
2570 SLSIASSESEKAKSEDEKRVNLSIGT--KQSKENOVSAKGTWRKIKENEFSP----- 2621
QY 2368 TSLQIPSGQPOQPOVQVOSTOTLSSGOT 2396
2622 -----TNSTSQTVSSGAT 2634

RESULT 7
US-08-289-548A-7
Sequence 7, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEO ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 2.4%; Score 362.5; DB 1; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8.1e-13;
Matches 409; Conservative 327; Mismatches 86; Indels 777; Gaps 98;

QY 373 WEAEIKLLEEMREIHHMDITDITLKARGSNKSFLLAANEELIESIRAKGIDIDNVK 432
D 698 WDMGAVSLKMLHSHKMI-----AMGS-----AALLRLMAMRAPKAYDA-NIM 742
QY 433 SPEET-----EKDKN-ETENDSKDAEKNEFEEDQSLKESDDKTPDDPEQKSEPT 485
D 743 SPGSSLPSELHVRKQKALAEALDAQHLETFDNIPLSPKASHRSKORHKOSLYGDYFDT 802
QY 486 EVGDKGNSVSNLGD-----NTT-----NATSEETSPSGRSPVGC--- 521
D 803 NRHIDNRSDNNTGMVLSPLYLNTVLPSSSSSKGSLDSRSREKDSLENERGIGLGN 862
QY 522 --LSETPDSSMAEKVASELPQ--DYEEPEN--KTCSSNTSATTTISI-----QPNLEN 570
D 863 HPATEPQTSSKRGIIQITTAQIAKVMEEVSAIHTSQEDRSQSGSTTEHCVTERNALR 922
QY 571 SNSSSELNSSOSEKAKADDPENGEBSHRTVYSIOELVGDFTSEKSTGELSESPGAGK 630
D 923 RSSAAHTSHNTYNTKTS---ENSNRFCSMFYA---KLEYRRSSNDLSNSSSDGYGKR 975
QY 631 ASGSTRIITRLRNPDSKLSQLKS--QOYAAAHEANKL-FKEGKEVLVYN----- 677
D 976 GQMPSTISYSEDDSEKSCSYQYLPADLAHKIHSANHMDNDGELDPIINSLKYSDEQL 1035
QY 678 SOGELSLSTKEVIMKGINNRYKLGQEGRYVHNOYSTNSPALNKHQREDHK-RR 736
D 1036 NSGQSPQONERMAPKHIIDEIKQSPQRSR--NOSTT--YPVYTESYDDKHLKQFP 1090
QY 737 HLAHFKLITPAGFEFKMNSVHGSKVLITSLRLITTOLENNIPSEFHPNMASTRANWIK 796
D 1091 HFGQOECVSPYRSRGANS-----ETNRVGS-----NHGINQNV 1124
QY 797 AVQMSKPRREFALALALIECAVRVYMLPIWREFLGHTRLHRMTSIEREKEKVKKEK 856
D 1125 SQSLCQED-----DYEDOKPTNYSERY 1146
QY 857 QEEBEIMQOATWVKYTPPVKHQVQKQEGEYRVTCYGQMSIKTHVYRPVPLRGNTNV 916
D 1147 SEEE-----OHEEERPTNY-----SI 1163
QY 917 NYRKSLEGTNNMNDENDES-----DKRCSPPKIKIEPDSEKDEV 959
D 1164 KYNEE-----KHHVQPIIDYSLKATYDIPSSOKOSFSFSKSSSGSTKEHHSSSEMTST 1219
QY 960 KGSDAAGADONEMDISKITEKKDQDYKELDSDS-----DKP-CKEEPMEV 1005
D 1220 PSSNAKR---QNLHPSSAQRSGQPOKATCKYSSINOETIQYVCVEDPDPICSRSSSL 1276
QY 1006 DDDKKTSHVNCQESSQ-----VDVYVNSEGFHLRT-----SYKKKTKSKSL 1047
D 1277 SSLSSADELIGCNOTQOADSANTLOJLAIEIKELGTRSAEDPVEVPAVSQHPRTKSRSL 1336
QY 1048 DGLLEKIKQFTLEKQRLKLEIGIKIGKSTNSSKN-----LSESPVT-TAKKEG 1101
D 1337 QG-----SSLSESARHNAVEFSSGAKSPSKGAQTTPKSPREHYVQETPLMFSRCTSY 1389
QY 1102 COSPMQOEQSPNANNQOPEDLLQGCOSDSVYLRMSDPGHTTKLTKPKRVLDDVIRS 1161
D 1390 SSLDSEFSRSLIASSVQSEP-----CGMNVSGIISPSDLDPDSQOIMPPSR----- 1434

QY 1162 PETCKPQNSIENDIEEKVSDLASRQEPPTKTKTGNDFIDSKLASAD----- 1211
D 1435 SKTPEPPPTQATQRTRE-----VPKNKAPTAKERESG-----PKQAAVNAAVQRVQVLP 1482
QY 1212 DIGILL-----CKKK-----KPLIOESDITV-----SSSKALHSSVPK 1246
D 1483 DADTLHFAETESTPDGFSKSSLSALSLDEPFIQKVDLRIIMPVQENDNNGNETESQPK 1542
QY 1247 STNDRDATPLSRAMDFEKLGCDESENSTLENSSDTYSDIOSSEEDMIYONSNSISQF 1306
D 1543 ESNENQ-----EKEAETKIDSEQ--LLDDSDDD----- 1569
QY 1307 RTRQDVEVLEPLKCELVSGESTNCEDRLPVKGTENGKKRQOQKLEERPYNKCSQI 1366
D 1570 -----DIEILE-BC-IISA-----MPTKSSR-KAKPAQTASKLPPVAKRKSQ 1611
QY 1367 ---KLKNTDKKNNENRESEKKQRTSTFOJNGDKNPKYILKGECLKEISESRVSGNV 1423
D 1612 PVYKRLPSONRLOPKHVSFTPG-----DDMPRVY---CV---EGTPIMFST 1652
QY 1424 EPKVNINIKIIPENDIKSLVKEGAIPEPTINGDVIMEDFNERNSSEKSHLLSSDAEGN 1483
D 1653 ATSLSDLTIESPPNEL-----AAGEVYGAQSGEFKEKRTIIPTEGR--STDEAQGG 1702
QY 1484 YRDSLETLPTSKES-----DSTQTTTPSACSES 1512
D 1703 KTSVS-PIPELDNKAEBGDIACINSAMPKSKSHKPFVRYKIMDQVQASASSAPNK 1761
QY 1513 NSVQVEDMEIETSEVKKYS-----SPITSEESNLNDP 1548
D 1762 NQLDGKR--KKPTVPKPIPONTEYRTVRKNADSKNNLAERYVSDNKSKONLKNNS 1819
QY 1549 IDENGCLPINKENVNGE-----SKRKTY 1571
D 1820 KDFNDKLPNNEDRKGSFAFSPHHTYPIEGTPICFSEFRNDSLSLDDDDVDLSRKK-- 1877
QY 1572 ITVEVTTSTVATESKTVIKVEKGQOTVYVSTENCASKSTVT-----TTTITVTK 1621
D 1878 -AELRAKAKENKESBAKVTSHTELSNQOASANKQAIKQPIINRQPKPILOKOSTPQSS 1936
QY 1622 LSTPSTGSDVDIIVKQOSKTVITTYTDSLITTTGGGLVTSMTYSKYSTROKVKLMKFS 1681
D 1937 KDIPDRGAAND--EKQLQNAIENTPVCFS--HNSLSLSLSDIDQNNKNEPIKETE 1990
QY 1682 RP-----KTRSGTALPSYRK-----FVTKTSKSIPLVLPDDP-----K 1716
D 1991 PRDSQGEPSKRPQASGYAPKSFHVEDTPVCFSRNSSLSSLSIDSEDLLOECISAMPKK 2050
QY 1717 KLARKGIREVPYFYNNAKP-----ALDIWYPSRPRTFGIT-----WRYRL 1758
D 2051 KPSRLKGDNE-----KHSPRNMGGILGEDLTLDKIQIPDSEHGLSPSENMDFWKAIQ 2104
QY 1759 QTVKSLAGVSLMLRLMASLRMDMAKVPGGSGSTTETSHSEITITTELIKRDVOPYG 1818
D 2105 EGANSIV-----SSLHQAAAACL-----SKQASDSD--SLSLKS--G 2140
QY 1819 IREFYCIKILICPGVETKETPTPQKGLRSALPCKRPPEPKQGPYIETFWAAEE 1878
D 2141 I-----SIGSPHLLTPDDEKPFISNKPRL--ILKPGKST-----LETKLIESE 2183
QY 1879 -----LEMEIRAFARVEKEKAQAVEQAKRLBOQR-----TVIATST 1919
D 2184 SKGIGKGAKEYKSLITGKVRNSE-----ISQGMQPLDANMPISRGRTMHIHG 2234
QY 1920 TSPSSSTTISPAQKVMAPISGSVTTGTKMVLTTKVSFPATVTTQOKNFHQRTATWY 1979
D 2235 VRNSSSTSTSPVSKGPKLPKTPASKSPSEBQGTATTTPSGAKPVY----- 2277
QY 1980 KQGSNGGVYQOQKVILGIIIPSSGTGSOQFTTFQOPTATVTRIPMTSGSGGTTNSQVY 2039
D 2278 ---KSELSPVAKQTSQIG-----GSK-----APSRSGSRDSTPS----- 2309


```
TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2842 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: protein
:   ORIGINAL SOURCE:
:     ORGANISM: Homo sapiens
:   IMMEDIATE SOURCE:
:     CLONE: APC
:   SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match      2.4%; Score 362.5; DB 4; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

Oy 373 WEAEKLIIEERREIHRMIDTETLTNKGSGNSKFLAAAEIESTIRAKKGIDIDNYK 432
Db 698 WDMGAVSMKNIHSHKHNI-----AMGS-----AALRLNLMANPAKTKDA-NIM 742
Oy 433 SPEET-----EKDN-ETENDSKDAEKNRFEEDOSLEKSDDDKTPDDPEOGKSEPT 485
Db 743 SPGSSLSPLSHVKKOKALELDAQHLSEFTFNIDNLSPKASHRSKORHKOSLYGDVFT 802
Oy 486 EYGDGKGNYSANLGD-----NTT-----NAISEETSPREGSPVC----- 521
Db 803 NRHDNRSDNFTGMNTVLSPLNTVLPSSSSSGSLDSSSEKDRSLERERGIGLVN 862
Oy 522 --LSETPSSNAEKVASELPQ--DVPEPN--KTCESSNTSATTTSI-----QPNLEN 570
Db 863 HPATENPGSSKRGIOISITTAQIAKMEVSAIHTSQDRSGSTTELHCYTDENALR 922
Oy 571 SNSSSELNLSGSESAKADDPENGRESHTPVSIQEIYDFTSEKSTGELSESPGAGK 630
Db 922 RSSAHTSHNTNYFTKS-----ENSNRTCSMPYA--KLEYKRSSNDLSNVSSSDGYGR 975
Oy 631 ASGSRITIRLNPDSKLSQLKS--QVAAAHAEANKL-FKRGKEVLVN----- 677
Db 976 GOMKPSIESYSEDDESKECISYQYPAADLAHKIHSANHMDDNGEDLPTPIYNSLYSDEQL 1035
Oy 678 SOGEISRLSTKEKVEIMKGINNYFKLOGEKYRVYHNOYSTNSFALNKQHREPDHK--BR 736
Db 1036 NSGRQSPSPNEMKARKIIEDEIKOSDQROSR--NOSTT--YPYTTSTDDKHLKFP 1090
Oy 737 HLAHFCULPAGEFKWNGSVHGSKVLTISTLRITLTLENNIPSSFHPNMASHRAMWK 796
Db 1091 HFGQOECEVSPYSRGANGS-----ETNRVGS-----NHGINQNV 1124
Oy 797 AVOKSKRPEFLALALIECAVAPVVMPLIRKEFLGHTRLHMTSLIEREKEKVKKKEK 856
Db 1125 SOSLQOED-----DYEDDKPTNYSERY 1146
Oy 857 QGEETMOATVWYKTPYKHOVMKQGEYRVYTGSGWSISKTHVYFVFKLPENTNV 916
Db 1147 SEEE-----OHEEERPTNY-----SI 1163
Oy 917 NYRKSLEGTKNMNDENMDS-----DKRKSRSFKKIKIEPDSKEDV 959
Db 1164 KYNE-----KRHVDPIDYSLKYATDIPSSOKOSFSFKSSSGGSKTTHMSSSSTNTST 1219
Oy 960 KGSDDAKGADQWEMDISKTEKKDDVKELLDSDS-----DKP-CKEPEMEV 1005
Db 1220 PSSNAKR--NQMLHPSSAQSRSGOPQKATCKVSSINOETIOTYCEVEDTPICFSCSSL 1276
Oy 1006 DDDMKTESHVNOCSQO-----VDVNVVSGFHLRT-----SYKKTKTSKRL 1047
Db 1277 SSLSSAEDEIGCNOTTQEDASNTLIQIAIKKIKGRSAEDVSEVPANVSQHPRTKSSRL 1336
```

```
Oy 1048 DGLERRIKQFLEEKORLEKIKLEGIGICKTSSNSKN-----LSSEPVY-TRAKEG 1101
Db 1337 QG-----SSLSESAHRKAVESGSAKSPSKSGQOTKSPREHYVQETPLMFSTCTSV 1389
Oy 1102 COSDSMRQOSPRANNDQEDDLQGCOSODSSVLRMSDPSHTNKKLYPPDRVLDDVSI 1161
Db 1390 SSLDSFESRSIASSVOSEP-----CSGAVSGIISPSDLDPDSGQTMPSR----- 1434
Oy 1162 PETKCPKQNSIENDIEEKYSDLASRGOEPTKSKTKGNDFIDSKLASAD----- 1211
Db 1435 SKTPPPPTQATQKR-----VPKNAKPTAEKREG-----PKAAVNAVAVQRYVLP 1482
Oy 1212 DIGTLI-----CKNK-----KPLIOESDPTV-----SSKSALHSSVPK 1246
Db 1483 DADTLHFATETSPDGFSCSSSLALSILDEPFIOKDVLEIRIMPPVQENDNGNETESEQPK 1542
Oy 1247 STNDRATPLSRAMDEGKLGCDSESNSTLENSDPVSIQDSSEEMIVQNSNESTSEQF 1306
Db 1543 ESNENG-----EKEAKTIDSEKD--LLDSDDD----- 1569
Oy 1307 RTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTAEANKKPPSOOKLEERPYNKCSQOI 1366
Db 1570 ----DIEILE-EC-IISA-----MPTKSSR-KAKKPAQTASKLPPYARKRPQL 1611
Oy 1367 ---KLNNTDKKNENRESEKKGQRTSTFOINGKONKPKIYLKECLKETSISRVSQNV 1423
Db 1612 PVYKLLPSQNRLOPKVHVSFTPG-----DDMPRYV--CV-----EGTPINFST 1652
Oy 1424 EPKVVNNINKIIPENDIKSLTVKESAIRPFIINGVIMEDENERSSTFKSHLSSDAEN 1483
Db 1653 AVSLSDLTIESPPNEL-----AAGEVGGAQSGEPEKRTJITEGR--STDAQGG 1702
Oy 1484 YRDLTLELSTKES-----DSTQTTTPSPASCPS 1512
Db 1703 KTSVV-TIEBLDNKAKBEDDILAECSINAMPKSKHKPRVKKIMQVQAASASSAPNK 1761
Oy 1513 NSVNOVEDMEIETSEVKKYTS-----SPTSEESNLSNDF 1548
Db 1762 NOLDGRRK--KKPTSVPKPIPONTEYRTVRVKNADSKNNLNAERVFSDNRSKQNKNN 1819
Oy 1549 IDENGLPIKNENVNGE-----SKRKTIV 1571
Db 1820 KQFNDRKLPNMEDVVRGSPAFDPSPHHTYPIEGTPYCFSRNDSLSLDPDDDDVLSREK-- 1877
Oy 1572 ITEVTMTSTVAETSKTVLKEGDKQTVVSSPTENAKSTVT-----TTTTPYVK 1621
Db 1878 -AELRAKENKESAEAVVTSHTELTSNOOSANKQALAKQPIINGOKPPILOKOSTPOSS 1936
Oy 1622 LSTPSTGGSVDIISVKEQSKTVVTTVTDLSLTTGGTLVTSMTVSKEYSTRDKVKLMKFS 1681
Db 1937 KDIPDRGAATD--EKLOMPAIENTVPCS--HNSLSLSLSDIDQENNNKENEPIKETE 1990
Oy 1682 RP-----KTRRSALSYRK-----FYTKSTKKSJFVLBNDL-----K 1716
Db 1991 PPDQSEPKRPAASGYAPSFHVEDTPVCFSRNSSLSSLSISEDDLLOECISSAMPKK 2050
Oy 1717 KLARKGIREVPEFYNNAPK-----ALDIIPYSPRPTFGIT-----WRYRL 1758
Db 2051 KPSRLKGDN-----KHSFRNMGILGEDLTLDLDIQRPSEKLSIDPSENFPMKALQ 2104
Oy 1759 QTVKSLAGVSLMLRLMLASLRMDMAKAVPPGGSGSTRFTSETETITTIKRDVGPYG 1818
Db 2105 EGANSIV-----SLHQAAAACL-----SRQASDSD-----SILSKS---G 2140
Oy 1819 IREYICIRKIICPGVPEPKETPTQKGLRSALRPKRPETPKOTGVYIITWAEER 1878
Db 2141 I-----SLGSPFHLTPDOEEKPFTSNKGP--ILKPEKST-----LETKRIESE 2183
Oy 1879 -----LELMERAFERVEKBEKAQAVEQAKKRLBEOQK-----TVIATST 1919
Db 2184 SKGIGKGRKVKYSLITGKRNSN-----ISOMQPOLAANPSPISRGTMTHIG 2234
Oy 1920 TSPSTSTSTISPAOKVWVAVPISGSVTTGTGKMYLTTVKGVSPATVTFQONKNFHQGFATV 1979
```

```

Db 2235 VRNSSSTSPVSKKPPPLTPPAKSPSECOATTPSPGAKPSV----- 2277
QY 1980 KOGOSNGVVOYQOKVLIIPSSSTGTSOQTFSTFOPTATVTRPMTSGSGGTTNSOVI 2039
Db 2278 ---KSLSPARQTSOIG-----GSSK-----APSRSGSRDSTPS----- 2309
QY 2040 TGPQIRPGMTVITPPLQOSTLKGATITRTPYMWPQ-----APQOVMTQIIRGQPVSTAVSA 2095
Db 2310 -----RPAQOPLSRPIQSPGRNS-----ISPGRNGISPPKRLSOLPRTSSPTA--- 2353
QY 2096 PNTVSTPPOKSLTS-----ATSTSNIOSSASOPPPQOGVYKLTMAOLTQITQGH 2146
Db 2354 -STKSSGSGKMTSTSPGRQMSQONLTKQICLSKNASSIRSEAS-----KGLNQNNNG 2407
QY 2147 GGNQGLTVLIQGGQTTGOLQ-----LIPQVTVLPBGQOLQMAAMPNGTVQRFLE 2199
Db 2408 GANKKVELSHMSSTKSSGSDRSERPVLYRQSTFIKEAPSPILRRKLEESAS-----FE 2462
QY 2200 PLATTTATTTTSTTSTAAGTGBQROSLSP-----QMQVHQD-----KTLPPAQS 2248
Db 2463 SLSPSSRPAPSTKSOQTPV-----LSPLDPMISLTHSSVOAGGWRKLPPLSP 2512
QY 2249 SVGPAAQOPOTAPORPOTOPQSPAPQEPVOTOTVSSHVSEAPTHAQS 2308
Db 2513 TIEYNGRPARKRDIAHSHESPLPINSGTWKRE---HSKHSSLPVSTWRRTGSS 2369
QY 2309 KPOVAOSQPOSNVQOGSPYRVOVSPSOTRIRPSTPSOLS-PGQOQOVQITTSQPIPIQH 2367
Db 2570 SSILSSSESESEKAKSDEKHVNSIGT--KQSKENOVSAKGTWRKIKENEFSF----- 2621
QY 2368 TSLQIPSGQPOSGPOVOSTOTLSSGOT 2396
Db 2622 -----TNSTSQTVSSGAT 2634

```

RESULT 10
US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107, 49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-452-655B-2

Query Match 2.4%; Score 362.5; DB 1; Length 2843;
Best local similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

QY 373 WENELCKIIEEMKEELHHRMDITEDLTNKARGSNKSFIAANBELLESIRAKKGIDNVK 432
Db 699 WDMGAVSMLKNIHSHKMI-----AMGS-----AALRLNLANPAPAYKKA-NIM 743
QY 433 SPEET-----EKDN-ETENDSKDAENREEDQSLERKSDDKTPDDDPQKSEPT 485
Db 744 SPQSSLPPLHVRKQKLEAELDQHLSETFDNDNLSPKASHRSQKRNKQSLYGVDFDT 803
QY 486 EVGDKNVSVANIGD-----NTT-----NANSEETSPBEGSPVGC--- 521
Db 804 NRHDDNRSDNFTNGNMTVLSPLYLNTVLPSSSSSRGSLDSREKORSLEBRGIGLQNY 863
QY 522 --LSETPDSSNMAKKVYASLPO--DYPEEPN--KTCSSNTSATTTSI-----OPNLEN 570
Db 864 HPATENGTSSKRGLOISTAAQIAKYMEEVAIHTSQEDRSSGTTLEHCYTDENALR 923
QY 571 SNSSSEINSSQSESAKAADDPENGERSHTPVISOEIVGDTSEKSTGELSESPGAGK 630
Db 924 RSSAHTHSNTYNTFS---ENSNRICSMFYA---KLEYKRSSNDLSNVSSSDYGKR 976
QY 631 ASGSTRITLRNPDKSLSLKS--QQVAAAHEANKL-FKEGEKVLVN----- 677
Db 977 GQMKPSTSESYDEDESFCSYGYQPADLAKITSAHMDNDDELTPPIVSLKYSDEOL 1036
QY 678 SOGEISRLSTKEVIMKGINNYFKLQOEGKYRVYHNOYSTNSFALNKHQREDHDK-RR 736
Db 1037 NSGRQSPSONERNARPHIIEDEIKOSEQRQSR--NQSTT--YPYTESTDDBKHLKQOP 1091
QY 737 HIAHKFCLPAGFEKNGSVHSGKVLITLTLTLTQLENNITSSFFHPMAHRAKNWK 796
Db 1092 HFCQOECVSPYRSRGANGS-----ETNRVGS-----NHGINOVY 1125
QY 797 AYOMCSRPREFALALILECAVPVWMLPIWREFLGHTRLHRVTSIEREKEKVKKKKK 856
Db 1126 SOSLQOED-----DYEDDKPTNSERK 1147
QY 857 QEEETWQATVWKYTFPVKHQYWKQGEERYVYTGGMWSKITHVYRFPKLPONTNY 916
Db 1148 SEEB-----QHEEERPTNY-----SI 1164
QY 917 NYRKSLEGTNNMNDENDES-----DKKGRSPFKKIKIEPDSKDEY 959
Db 1165 KYNE-----KRHYDQPIDYSLKATYDIPSSQKOSFSFSKSSSQSKTEHMSSSSTNTST 1220
QY 960 KGSNAAGCAQONEMDKITEKKDQVKELLDSDS-----DKP-CKEPEMEY 1005
Db 1221 PSSNAKR---QONLHPSSAOSRSGQPOKAAATCYVSSINQETIOTTYCEVEDPPIFSCSSL 1277
QY 1006 DDDMKTESHVQCOSSQ-----VDVAVNVEGFHLRT-----SYKRTKSSKL 1047
Db 1278 SSLSAABDEIGCNGTTOFADASANTLQIAIEIKKIGTASADPVSEVPAVYQHRPTSSRL 1337

```

1048 DGLLEERRIOFLEEKORLEKIKLEGGIKIGKTSTNSSKN-----LSSESPVI-TRAKEG 1101
1338 OQ-----SSLSESEARHKAVEFSSGAKSPSKGAQOTPKSPENHYQETPLMFSTSV 1390
1102 CGSDSMRQOSNMANNNDOPEDLQGCSDSDSVLRKMSDPSSHNTNKLYPPRDVLDVYSIS 1161
1391 SSLDESSESNIASSVOSE-----CSGMSVSGIISPSDLPDSCGQMPSPR----- 1435
1162 PETKCPKONSINDIEEKKVSDLASRQCEPTKSTKGNDFIDSKLASAD----- 1211
1436 SKTPEPPPTAQTRE-----VPKKKAPTAKRESG-----PKQAAVNAQVQVLP 1483
1212 DIGTLI-----CKN-----KPLIOESDTIV-----SSKSALHSSVVK 1246
1484 DADTLHFATESTPDGFCSSSLALSALDEPFIQKDVLEIRIMPVQENDNGNTESEQPK 1543
1247 SUNDNDAFPLSAMDEEGKLGCDSESNSTLENSDPVSIQDSEEMITVQNSNESTSEOP 1306
1544 ESENEO-----EKEAEKTIIDSEKD-----LDDSDDD----- 1570
1307 RTREODVEVLEPLKCELVGSESTGNCEDRLPVKGTENGKRCPSQOKLEBERPVNKCSDOI 1366
1571 -----DIELE-EC-IISA-----MPTKSR-KAKKAQOTASKLPPYAKRPSOL 1612
1367 ---KLKNTTDKKNENRESEKKGQRTSTFQINGKDKPKITLKECLKEISERVYSGNV 1423
1613 PVYKLLPSQNRLOPKHVSFTPG-----DDMPRY-----CV-----EGTPINFST 1653
1424 EKVANNINKITPENIKSLTVKESAIRPINDVIMEDNERNRSTKSHLSSSABEGN 1483
1654 ATSLSDLTIESPNEI-----AAGEVKGAGSGEPEKEDTPTER--STDAQOG 1703
1484 YRDSLETLPTKES-----DSTOTTPPSACPS 1512
1704 KSSV-TIEPDDNKAEECDIIAECINSMAPKSHKPPRYAKKINDQVQASSSAPAK 1762
1513 NSVNOVEDMEIETSEVKVYS-----SPITSEESNLNDP 1548
1763 NOLDKCK--KKPTSPVKPIQPTXEYRTRVKRNADSKNNLNAERVFSDNKKONLKNNS 1820
1549 IDENGLPIKKNENVGE-----SKRKIV 1571
1821 KPFNKLPPNEDRVKGSFAEDSPHHYTPLEGPPCYCSRNDLSLDFDDDDVLSREK-- 1878
1572 IYEVTTMTSTVATESKTVIKVEKDKQVVSSTENCASKTV-----TTTTVTK 1621
1879 -AELKRAKENKSEAKVYSHTELTISQOSANKTOAIKAPINRGOPKPILOKOSTPQSS 1937
1622 LSTPSTGSGVDIISVKEOSKTVVTTVYDSLTGTGTLVTSMTVSKEYSTRDKVKLMKFS 1681
1938 KDIPORGAATD--EKLQFAIENTPVCS--HNSLSLSLSDIDQENNNKENEPIKETE 1991
1682 RP-----KTRSGTALPSTYK-----FYKSKTKSTFVLPNDL-----K 1716
1992 PPDSCGSEPSKQASGAPRSFHEVEDTPVCFSRNSSLISLSEDDLQOECISAMPKPK 2051
1717 KLARKGIREVYFYVNAKP-----ALDIMPYSPRPRTGIT-----WRYRL 1758
2052 KPSRLKQDNE-----KSPRMGGLGEDLTLDKDIDORPSEHGLSDSENFDMKAQ 2105
1759 QTVKSLAGVSLMLRLMLMSLRMDMAKVPYGGGSTRTETSETITTTTIIIRRDVPGY 1818
2106 EGANSIV-----SLHOAAAAACL-----SRQSSDSD-----SILSKS-----G 2141
1819 IRFECEIKIICPIGVPEPKPTPTPQKGLSSALRPKRPPTPKOTGVIIETVAAEE 1878
2142 I-----SLGSPHLTTPQEEKPPTSNKGP--ILKPEKST-----LETKKIESE 2184
1879 -----LELMEIRAFERKEKAQAVEQOAKRLEQOKP-----TVIATST 1919
2185 SKGIGGKKVYKSLITGKVRNS-----ISGQMKOPIQANMPSISRGRTHIHFG 2235

1920 TSPSTSTTISPAQVYMAVIGSVYTTGKMYLTKKVGSPATVTOQKNPHOTATWV 1979
2236 VRNSSSTSPVSKKGPPLTPAKSPSECOATATSPGAKPSV----- 2278
1980 KQOSNSGVVQVQKVLGIIPSSGTSGQFTSFQRTATVTRPMTVSGSGGTSMSOVI 2039
2279 ---KSELSPAROTSQIG-----GSSK-----APSRSGRSTPS----- 2310
2040 TGPQIRPGMTVTRPQOOSTIGKAIIRTPVMVQPG-----APOQVMQIIRGQPVSTAVSA 2095
2311 -----RPAQOPLSRPIQSGRNS-----ISGRNGISPPMKLSQLPTSSPSRA----- 2354
2096 PNTVSTPGQKSLTS-----ATSTNSIQSSASQPPRPGQGVYKLTMAQITQLTQGH 2146
2355 -STKSSGSKMSYSTSPGRMSQONLTKQTLGSKNASSIPRESAS-----KGLQNMNGN 2408
2147 GGNQGLTVIYQGGQOTTGQLO-----LIPQVTVLPGQOOLMAAMPNGVQVFLFT 2199
2409 GAKKVELRMSSTKSSGSDRSRERPVLRQSTFIKEAPSPTLRRKLEESAS-----FE 2463
2200 PLATVATTAATTTTSTTAAGTGEOQRQSLSP-----QMOVHOD-----KITLPPAQS 2248
2464 SLSPSSRPASPTSQOTIV-----LSPLDMSLSTHSSVQAGMKRLPRLNLS 2513
2249 SVGPANAQPOTAQPSARPOPOTQOPQSPAQPEVOTQOTVTSVHPSEAOPTHAQS 2308
2514 TIEYNDGRPAKRHDIAHSSEPSRLPINSGTWKRE--HSKSSSLPVTSTWRRTGSS 2570
2309 KPOVAAQOSQVQOSQSPRVQSPQOTRIRPSTPQLS--PGQOSQVOTTTSPITQIPH 2367
2571 SSTLSASSESEKAKSEDEKHVNSISGT--KQSKENOVSAKGTWRKIKENEFSP----- 2622
2368 TSLQIPQOGQPOQOPOVOSTOTFLSSGQF 2396
2623 -----TNSTSQTVSSGAT 2635

RESULT 11
US-08-452-655B-7
; Sequence 7, Application us/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: us/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548

FILED DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ANTI-SENSE: NO
US-08-452-655B-7

Query Match 2.4%: Score 362.5; DB 1; Length 2843;

Best Local Similarity 17.3%; Pred. No. 8.1e-13; Indels 777; Gaps 98;
Matches 409; Conservative 327; Mismatches 856;

QY 373 WEALCKILEMEBEIHRMDITEDLTNRKSGNSKSFLLAANEELIESIRAKKGDIDNVK 432
DB 699 WDMGAVSMLKLNLIHSKHMI-----AMGS-----AALNLNLANRAKTKDA-NIM 743
QY 433 SPREI-----EKDKN-ETENDSKDAEKNEEEDOSLEKSDDKTPDDDEGKSEPT 485
DB 744 SPSSSLPSLHVROKALELDAQHLETFDNIIDNLSPKASHRSKORHKOSLGDVYVDT 803
QY 486 EVDGKGSYSANLGD-----NTT-----NATSEETSPSGRSPVGC--- 521
DB 804 NRHDDNRSDNFNTGNMTVLSPLYNTIYVLPSSSSSGSLDSRSREKDRSLEREKIGLGNV 863
QY 522 --LSETPDSSNMAEKVASELPQ--DYPEEPN--KTCESSNTSATTTST-----QPNLEN 570
DB 864 HPATEPCTSGSKRGLOISTTAQIAKYMEEVSAIHTSQEDRSGSSTELCHCVDERNALR 923
QY 571 SNSSSELNSQSASAKAADPENGGERESHPTVSIQEIYODTSEKSTGELSSPAGKG 630
DB 924 RSSAAHTHSNTYFTKS---ENSNTCSMPYA---KLEYKRSSNDLSVSSSDGYGR 976
QY 631 ASGSTRIILRLRPDSKLSQLSK--QOVAHAAHEANKL-FKEGEVLYVW----- 677
DB 977 GQMKPSLIESYSEDESKPCSYGYQYPADLAKIHSAHHMDNDGDELTPINYSIKYSDEOL 1036
QY 678 SQGEIIRLSTKKEVIMKGINNMFYKLGQEGKYRVYHNOYSTNSFALNKHQHREDHDK-RR 736
DB 1037 NSGRQSPSONERWARKHIEIEIKOSQROSR--NOSTT--YPYTESTSDKHFKQP 1091
QY 737 HLAHKFCLTPAGFEKWNQSHGSKVLITSLRTITQLENNIPSPFFHNMASHRANWLK 796
DB 1092 HREGQGCVCVPIYRSGANGS-----ETNRVGS-----NHGINQV 1125
QY 797 AVOMCSKPREFALALAILCAVAVVMLPIWREFLCHTRLHMTSLIEREKEKVKKEKK 856
DB 1126 SOGLCOED-----DYDDDKPTNYSERY 1147
QY 857 QEEELFMQATVWKTTPPVKHOVMKQGEERYVYTGGSWISKTHVRFVKLPDNTNV 916
DB 1148 SEEE-----OHEEEERPTNY-----SI 1164
QY 917 NYRKSLEGTKNNDENMDES-----DKRKSRSPPKKIKIEPDEKDEV 959
DB 1165 KNEE-----KRVHDDPITDLSKAYADIPSSQKOSFSSKSSGSSKTEHMSSSSNTST 1220
QY 960 KSSDAKAGADONEMDISKITEKKDQVKEILDSDS-----DKP--CKEEMPEV 1005
DB 1221 PSSNAK---QNLHPSSASQSRSGQPOKAATCKVSSINQETIOTYCVEDTPICFGRCSL 1277

QY 1006 DDDMKTFESHVNCQESSQ-----VDVNVNSEGFHLRT-----SYKKTKSSKL 1047
DB 1278 SSLSASADEIGCNOTQOADSANTLQIAETKEIKIGTSABDPVSEVPAYVQHPRTSSRI 1337
QY 1048 DGLERRIKOFTLEEKORLEKIKLEGIKIGTSTNSKN-----LSESPVI--TAKREG 1101
DB 1338 QG-----SSLSSEKRRHAAVEFSSGAKSPSKSGAOTPKSPPEHNVQETPLMFSCSTV 1390
QY 1102 QOSDSMRQEOSPANANDQPEDDLIOGCSQSDSVLRMSDPSHTTNKLYPKDRLVDVIRS 1161
DB 1391 SSLDSEFSRSIASSVQSEP-----CSGWVSGITISDLPDPSGQIMPSPSR----- 1435
QY 1162 PETCKPKONSIENDIEKVSADLASRGOEPKSKTKGNDPFIIDSKLASD----- 1211
DB 1436 SKTPPEPPPOAQTQKRE-----VPKNAAPAEKRESG-----PKOAAVNAVOROYVLP 1483
QY 1212 DIGTLI-----CKNR-----KPLIOESDPTIV-----SSKSALHSSVPR 1246
DB 1484 DADTLHLHFAFESTPDGFRSCSSLSALSALDPEFTQKDVLELRIMPVQENDNGNTESEQPK 1543
QY 1247 STNDRATPLSRAMDCEGKICGDSSENSLIENSSTVSIQDSSEDMIVQNSNESTSEOF 1306
DB 1544 ESNENO-----EKEAKTIDSEKD---LIDSDDD----- 1570
QY 1307 RTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTANGKPKPOQKLEBRPVNKCSDOI 1366
DB 1571 -----DIEILE--EC-IISA-----MPTKSSR--KAKKPAQTASKLPPYARAKPSOL 1612
QY 1367 --KLKNTYDKKNENRESEKKQORTSTFOINGKDNPKIYLKCELKEISESRVVSNGV 1423
DB 1613 PVYKLLPSQNRLOPQKHVSFTTP-----DDMPRYV-----EGTPINFST 1653
QY 1424 EPKVNINIKIIPENDIKSLVKSALRPLFNGDVIMEDNERNSSFTKSHLSSDAEEN 1483
DB 1654 AUISLDTLIESPNEL-----AAGEVREGAQSGEFEKRTDPTBER--STDAQOG 1703
QY 1484 YRDSLETLPSTKES-----DSTQTTTPSASCPS 1512
DB 1704 KTSV--TPELDONKAEEDGTLAECINSAMPKGSKHPRVYKIMQVOAASASSAPRK 1762
QY 1513 NSVNOYEDMEIETSEVKKYTS-----SPITSSEBSLNSDF 1548
DB 1763 NOLDGKR--KKPTSPVAKPIPONTERTVRKNKADSKNLNAERVSNDKSKRONKNS 1820
QY 1549 IDENGLPIKNENRVNCE-----SKRKTV 1571
DB 1821 KDFNDKLPNNEDRVKSFAPDSBHHTPIEGTPYCFSRNDSLSLDPDDDDVDSLREK-- 1878
QY 1572 TTEVTMTSTVATESKTVIKERGDQIVVSSSTENCAKSTVT-----TTTTVYTK 1621
DB 1879 -AELRAKENKSEAKVTSHTELTSNOQSANKTOAIKQPIRNGQPKPILQKQSTFPQSS 1937
QY 1622 LSTPSTGSGVDIISVKEQSKYVTTVTYDLSLTGTGLVTSMTVSKESVTRKQVLYMFS 1681
DB 1938 KDIPOGAATD---EKIQFALENTPVCS---HNSSLISLSDIOENNNKENEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FVTKSTKSKSFVLPNDL-----K 1716
DB 1992 PPDGSGEPKPPASGAYAPKSFHVEDTPVCFSRNDSLSLSDISEDDLLOECISSAMPKK 2051
QY 1717 KLARKGGIREVPYFVYNAKP-----ALDIWPPSPRPTFGIT-----WRYRL 1758
DB 2052 KPSRLKGDNE-----KHSFRNWGILGBDLTLDKDIQRPDSEHGLSPDSENFWMKAIQ 2105
QY 1759 QTVKSLAGVSLMLRLMLASLRMDMAAKVPPGSGSTRITETSTETITTTIELIKRDVGYG 1818
DB 2106 EGANSIV-----SSLHQAAAACL-----SROASSDSD---SLSLKS---G 2141
QY 1819 IFFECIRKIIICPIGVPTPKETPTPKRKGSLASLRARPKRPTPKQGTQPVIIETWVAEE 1878
DB 2142 I-----SLGSPFHLTPDQEKPKFTSNKGR--ILKFGKST-----LETKIESE 2184

```

QY 1879 -----LELMETRAFAERVEKEKAQVNEQAKKRLDQKP-----TVIATST 1919
Db 2185 SGIKGGKVVYSLTNGKRSNSE-----ISQOMKOPLOANPPSISRGRTMIHIG 2235
QY 1920 TSPTSTSTISPAKQVNVAPISGVYTGKMYLTKVSPATVTPQNNFQOTATW 1979
Db 2236 VNRSSSTSPVSKGPKPLTPASKSPSEGTATTSRGAKPSV-----2278
QY 1980 KQGSNSGVVQVQKVLGIIPSTGTSTQOTFTSPORATVTRPNTSGSGGTTNSQV 2039
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APSRSGSDSTPS---- 2310
QY 2040 TGPQIRPGMTVTRPLQOSTLKAITRTVMQPG---APQOVMTQIIRGQVSTAVSA 2095
Db 2311 ---RPAQOPISRPISQSGRNS-----ISPGRNGISPPNKLSQLPRTSSPSTA--- 2354
QY 2096 PNTVSTPGOKSLTS-----ATSTNISASQPPRQGGQVKLMAQLTQLQGH 2146
Db 2355 -STKSSGSKMSTTSGRMSQONLTQGLSKNASSIPSESAS-----KGLNMNNGN 2408
QY 2147 GNGQGLTVVIOGGQGTGQLQ-----LIPQVTVLPGRQOLMAAMPNGTVQRFLEPT 2199
Db 2409 GANKVELSRMSTKSSGESDSESRPVLVROSTFIKAPSPTLRKLLESAS-----FE 2463
QY 2200 PLATATTTSTTTTSTTAAGTGEOROSKLSP-----QOVYHOD-----KTLPPAQSS 2248
Db 2464 SLSPSSRPASPRSQOQTV-----LSPLSPDMSLSTHSSVQAGGWRKLPNLSL 2513
QY 2249 SVGPAAKQPTQASARPOPTOPQSPAPQEVQTOPEVOTQTVSSHVSEAPTHAQS 2308
Db 2514 TLEYNGRAKAKHDIARSHSESPSLPINKSGTWKRE---HSHSSSLPVSSTWRTGSS 2370
QY 2309 KQVAAQSQPQSNVQGSFVQVQSPQTRIRPSTPQLS-PGQSQVQVTTQSPPIPIOPH 2367
Db 2571 SSILSSSESSSEKASEDKHVNISIGT--KQSKENQVSAKGTWKRKIKENEFPSP----- 2622
QY 2366 TSLQIPSQQPOSQVQVOSTQVTLSSGOT 2396
Db 2623 -----TNSTQTVSSGAT 2635

RESULT 12
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582

```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

Query Match 2.4%; Score 362.5; DB 3; Length 2843;
Best local Similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

QY 373 WEAECLKILEMBEETIHRHMDITEDLTNKAQSGNKSFLAANEETIESIRAKKGIDIDNYK 432
Db 699 WDMGAVSMKLNLIHSHKHM-----AMGS-----AAALRNLMANPAKYKKA-NIM 743
QY 433 SPEET-----EKDN-ETENDSKDAEKNREFEQDSLEKDSDDKPRDDPEGKSEEP 485
Db 744 SPQSSLPPLHVRKQKALELDAQHLSEFTDNTIDNLSPKASHRSKORHKOGLGYVFDI 803
QY 486 EVGDKNSVSNLGD-----NTT-----NATSEETSPSEGRSPVGC--- 521
Db 804 NRHDNRSDNENFTGNMTVLSPLYLNTVLPSSSSRGLSDSRSEKRLSREHIGLGNV 863
QY 522 ---LSETPDSSNNAEKVVAELPO--DYPEPN--KTCESSNTSATTTST-----QNLN 570
Db 864 HPATENPGRSSKRGLOISTTAQIAKVMEVSAIHTSOEDRSSGTTTELHCYTDENNAIR 923
QY 571 SNSSSELNSQSESAAADDPENGESRHTPVYSIQEIVGDFSEKSTGEISPPAGK 630
Db 924 RSSAHTHSNTVFTKS---ENSNTCSMPYA--KLETKRSSNDLSNVSSSDGYGR 976
QY 631 ASGSTRIITRLRNPDSKLSQLKS--QOVAAAHAEANKL-FKEGKVLVYN----- 677
Db 977 GQWKPSIESYSEDEDESKFCSYGGYPADLAHKIHSANHMDDONGELDPTIINSLKYSDEOL 1036
QY 678 SOGEISRLTKKEVKINNGINNTFKLGQEGKRVHINQYSTNSFALNKHQHRDHDK-RR 736
Db 1037 NSGRQSPQSONERWARPKHIIEDIKOSEQORSR--NOSTT--YPPVTESTDCKHLKFP 1091
QY 737 HLAHKCLTPAGFEKNGSVHOSKVLITSLRLTITQLENNIPSSFFHNMASHRANWIK 796
Db 1092 HFGQOECVSPYSRGANGS-----ETNRVGS-----NHGINOV 1125
QY 797 AVQMSKPREFALAIALCAVAVPVMLPIWREFLGHTRLHRTSIEREKERVKKKK 856
Db 1126 SOSLQGED-----DYDDDKPTNVSEHY 1147
QY 857 QEEETMQATVTKTTPVKHQVWKQGEERYTYGSGNSATSKTHVYTFVKRLGENTV 916
Db 1148 SEEE-----OHEEERPTNY-----SI 1164
QY 917 NYRKSLEGTKNMDEMDES-----DKRKSRSPPKKIKIIPDSEKDEV 959
Db 1165 KYNE-----KRHYDPIIDYSLKAYDAIDIPSOQOSFSFKSSSGQSSKTBHMSSEBNTST 1220

```

```

QY 960 KGSDAKAGADQONEMDISKITEKKDQVKELLDSDS-----DKP-CKEEMEV 1005
Db 1221 PSSNAR-----QNLHPSASQSRGQPOKATCQVSSINQETIQTYCEVDPICFSCSS 1277
QY 1006 DDDMKTESHVNCOESSQ-----VDVNVSEGFHLRT-----SYKKTKSSKL 1047
Db 1278 SSLSSADEDEIGCQOTQOEDSANTLQIAELKEKIGTRSADDPVEVPAYASQHRTKSSRL 1337
QY 1048 DGLLEBRKQFLEEKORLEKILEGGIKGIGKSTNSSKN-----LSESPV-TRAKED 1101
Db 1338 QG-----SSLSESAKRAKAVFESSGAKSPKSGAOTKSPREHVOEPLPMSKCTSV 1390
QY 1102 COSDSMKROESPANNDQPEDLLIQCSQSDSVLRMSDPSHTNKLKPKDRLVDVIRS 1161
Db 1391 SSLDSESRNIASSVOSEP-----CSGMVSGIISPSDLPDPSGQIMPPSR----- 1435
QY 1162 PFKCKRQNSIENDIEEKVSDLASRGOEPTKTKGNDPFIDSKLASAD----- 1211
Db 1436 SKTPPPPOQAOTKRE-----VPKNKAPTAERREG-----PKOAAVAAVQROVILP 1483
QY 1212 DIGTLI-----CKNK-----KPLIOESDITV-----SSSKALHSSVPK 1246
Db 1484 DATLLHFATESPPDGFSCSSLSALSDEPFIQKDELRIIMPVOENDNGNTESEQRK 1543
QY 1247 STNDRDATPLSRAMDEEGKLGCDSESNSTLENSSDTVSJDSSSEDMIVONSSESIEQF 1306
Db 1544 ESMENO-----EKEAKETIDSEKD-----LLDSDSD----- 1570
QY 1307 RTEQOVEVLEPIKCELVSGESTGNCEDRLPVGTEANGKPKSOQKLEPRPVNCKSDOI 1366
Db 1571 -----DIEILE-EC-IISA-----MPTKSSR-KAKPAQOTASKLPPPAKRPQSOL 1612
QY 1367 ---KLKMTTKKNNENRESEKKGORTSTFOINGKNPKRIYLGECLEISESERVSGNV 1423
Db 1613 PVYKILPSQRLOPQKHVSTFG-----DDMKRY-----CV-----EGTPIINFST 1653
QY 1424 EPVYNNINKIIPENDIKSLTVKESAIRPFIJNGVIMEDENERNSETKSHLLSSDAEGN 1483
Db 1654 ATSLSDTLTISPPNEL-----AAGEGVGGAQSGEFEKRDITPEGR--STDEQOG 1703
QY 1484 YRDLSTLEPTKES-----DSTQTTPTASQSPES 1512
Db 1704 KTGSSV-TIFELDNKKAEGDILAECLNSAMPKSKSHKPPVKKIMDOVQOASASSAPNK 1762
QY 1513 NSVNOVEDMEIETSEVKKVT-----SEPTSEESNLNDP 1548
Db 1763 NQDQKK--KKPTSPVKPIQNTETRYRVRKKNADSKNNLNAEVEFSDNKKSKONLKNS 1820
QY 1549 IDENGLPIKNNVNGE-----SKRKIV 1571
Db 1821 KDENDKLPNNEEDVRSFAFSPHHYPIEGTPECFSRNDSLSDLFDDDDVDLSREK- 1878
QY 1572 ITEVTMTSTVATESKTVIAVEKGQOTVASTENCACKSTVT-----TTTTVTK 1621
Db 1879 -AEIRAKAKEKSEKAYTSTELTSSNOOSANKQAIKQPIINGQKRPILQKOSTPQSS 1937
QY 1622 LSTPTSGGSVDIISVKEQSKTVYTTVTDLSITTTGTLVSMTVSKKEYSTRDKVLMKFS 1681
Db 1938 KDIPDRCAAND---EKLONPAIENTPVCFS---HNSLSLSLSDIDQNNKNEPIKETE 1991
QY 1682 RP-----KTRSGTALBETRK-----FVTSYTKKSTIVLNDL-----K 1716
Db 1992 PPSQSGEPSPQASGYAPKPFHVEDTPVCFSRNSSLSSLSIDSEDLQOCISSAMKPK 2051
QY 1717 KLARKGIRVPEFVFNARK-----ALDIMYPSPRPPTFGIT-----WRYL 1758
Db 2052 KPSRLKGDNE-----KHSRNMGGILGEDLTLDLKDIOKRPSEHGLSPSENFDMKAIQ 2105
QY 1759 QTVKSLAGVSLMLRLMLRMDMAKVPBGGSSTRTETSETEITTEITTIKRDVGPYG 1818
Db 2106 EGANSIV-----SSLHQAAAAACL-----SRQASDSD--SILSLKS---G 2141
QY 1819 IREYCIKRIICPIGVETETKETPTQORKGLRSSALRPKRPETPKQGPVITETWAEEL 1878

```

```

Db 2142 I-----SLGSEPHLTPPOEKEPFTSNKGR--ILKPEKSP-----LETKIESE 2184
QY 1879 -----LELMEIRAFEREREKEKAQAVEQAKRLQOQK-----TVIATST 1919
Db 2185 SKGIRKGGKVVYSLTGKRSNSE-----ISQOMQPLDAMPSTRGKTMHIHG 2235
QY 1920 TSPSTSTTISPAQKVMVAPISGVYTGKMYLTTKRVGSPATVTPQOKNKHQTEFATV 1979
Db 2236 VRNSSSTSPVSKGKGPPLTPAKSPSEGTATTPRGAKPSV----- 2278
QY 1980 KQGSNSGVYVOQKVLGILTPSSTGTSQOTFTSFQRTATVTIRPNTSGSGGTTNSQVI 2039
Db 2279 ---KSELSPYARQTSQIG-----GSSK-----APRSGRSDSTPS----- 2310
QY 2040 TGFQIRPMTVIRTPLOQSTFLGKALIRTPVMVOPG---APQOVMTQIRGQPVSTAIVA 2095
Db 2311 ---RPAQOPLSRPIQSPGRNS-----ISPGRNGISPRKLSQLRTSSPSTA--- 2354
QY 2096 PNTVSTPGQSKLTS-----ATSTSNIOSSASQPPRPOGOYKLTMAQLTQLTQGH 2146
Db 2355 -STKSSGSGKMYTSPGRQMSQONLTKQGLSKNASSIPRESAS-----KGLNQNNNG 2408
QY 2147 GGNQGLTVYIOGGQTTGQLQ-----LIPQGVTVLPBPGQOLMAAMPNGVQFLFT 2199
Db 2409 GANKKVELSRMSTKSSGSESDRSEPRVIVRQSTFTIKKASPTLRKRLLESAS---FE 2463
QY 2200 PLATATTAATTTTGTVSTAGTGEOQRKLSF-----OMQVHQD-----KTLPAQSS 2248
Db 2464 SLSPSRPASPTSOQATPV-----LSLSLDMLSLTHSSVQAGWKRLPRLNLSF 2513
QY 2249 SVGPAAQOPQAPASARPQOTQOPQSPADPEVOGTQTTVSHVSEAOPTHAQSS 2308
Db 2514 TIEYNDGRPAKRHDIAHSESPSLRPLNRKSGTWKRE---HSKHSSSLPRVSTWRRTGS 2570
QY 2309 KPOVAAQOPQSVQOQSPVRYQSPQOTRIRPSTPOLS-PGQOQVQVTTSTQPIQGH 2367
Db 2571 SSILSASSSESEKAKDEXHVNISIGT--KQSKENOVSAKGTWIKENESP----- 2622
QY 2368 TSLQIPQOGQPOQOPOVOSTQTLSSQOT 2396
Db 2623 -----TNSSTQTVSSGAT 2635

RESULT 13
US-08-450-582-7
; Sequence 7, Application us/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: MAKAMURA, YOSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

QY 1759 QYKSLAGVSLMLRLMRLMDMAKVPGGSGTETETSETETITTEIKRRDVPYG 1818
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2106 EGANSTV-----SSLHQAANAACL-----SRQASDSD-----STLSKS-----G 2141
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1819 IREYECIRKICIPGVEFKETPTPOKGLRSSALRPKRPETPKOTGPVIEFWAEE 1878
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2142 I-----SLGSPFHLPPDEEKPFISNKGPR--ILKPGKST-----LEKKKESE 2184
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1879 -----LEMEITAPAEERKEKKAQAVEQAARLEQKP-----TVIATST 1919
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2185 SKGIRGKKYKYSKILGKVRKNSB-----ISGQMKOPLQANMPSISRGFTMHIPG 2235
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1920 TSPSTSTSTISPAQAKMVAPISGSVTTGFKMVLTTKVSAPATVTFQONKNQOTFTWY 1979
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2236 VRSSSTSTSVSKGPKPLTPAKSPSEGTATTPRGAPSV----- 2278
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1980 KOGOSNGVVOOVOKVLIIPSTGTISOQFTSFOPRTATVTIRPNTSGSGGTTNSQVI 2039
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APSKSGSRDSTPS--- 2310
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2040 TGPQIRGMVITRPLQOSTLGKAIITPVWVOPG---APQVMTQITIGQPVSTAVSA 2095
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2311 ---RPAQPLSRPIQSPGRNS-----ISPGRNGISPPNKLSQLPRTSSPSTA--- 2354
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2096 PNTVSTTPGOKSLTS-----ATSTSNIOSSASQPPPOGOVYKLTMAQLTOLTQGH 2146
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2355 -STKSSSGSKMSTSPGRMSQONLTQOTGLSKNASSIRPSESAS-----KGLNOMNGN 2408
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2147 GGNQGLTVIIOGQOQTGQIQ-----LIPQGVTVLPFGQOIQMAAMPNGTVQRFLE 2199
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2409 GANKKVLISMSSTSKSSGSDSEREVLVROSTFIKEAPSPILRRKLESAS-----FE 2463
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2200 PLATTATTATTTTSTTSTAAGTGEORQKSLSP-----QMQVHQD-----KITLPAQSS 2248
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2464 SLSSSPAPSPITSOQOTPV-----LSPSLPMSISTHSSVQAGMKLPPLSP 2513
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2249 SVGPAAQOPOTAPSPARPOQOTOPSPAPQPEVOTQOTVTSHPSEAOPTHAQSS 2308
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2514 TIEYNDGRPAKRHDIAHSHSESPRLPINSGTWKRE---HSKSSSLPVPSTRRGSS 2570
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2309 KPOVAASQOSNVQOSPVAVOSPSOTRIRPSPOLS-PCQOSOVQOTTTSQPIPIQPH 2367
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2571 SSILSASSESEKAKSEDEKHNVISGCT--KQSKENQVSAKGTWRKIKENEFSP----- 2622
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2368 TSLQIPSOQOPQOSQOTOTLSSGOT 2396
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2623 -----TNSTSQTVSSGAT 2635
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,355A
FILING DATE: 20-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UT
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5851775e
US-08-821-355A-7

Query Match          2.4%; Score 362.5; DB 2; Length 2973;
Best Local Similarity 17.3%; Pred. No. 8,6e-13;
Matches 409; Conservative 327; Mismatches 866; Indels 777; Gaps 98;

QY 373 WEAFCKILEEMREERIRHMDITEDLTNKGARSKSFLAANEELIESIRAKKGDIDNVK 432
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 WDMCAVSMKLNLIHSHKIM-----AMGS-----AALNLMANRPAKTKDA-NIM 743
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 SPEET-----EKDN-ETENDSKDAEKNEFEEDOSLEKOSDDKTPDDPEOGKSEEP 485
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 744 SPGSLPLSLHVKKQKALELELDQHLSETFDNIDNLSPKASHRSKORHKOSLGDYVFD 803
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 EYDGKNGSVSANIGD-----NTT-----NATSETPSPSGRSPVC--- 521
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 804 NRHDNRSDNFTNGMTVLSPLYLNTVLPSSSSRGLSDSSRSKDRSLERENGIGLGN 863
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 --LSETPDSSNMAEKVASELPQ--DYVEEPN--KTCESSNTSATYTSI-----QPNLEN 570
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 864 HPATENGTSGSKGLQISTAAQIAKVMESAIHTSQEDRSSGSTTELHCYDERNALR 923
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 571 SNSSSELNQSQESAKAADDPENGERSHTPVSIQETVGDFTSEKSTGLSSPGAGKG 630
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 924 RSSAAHTHSNTYNTKTS---ENSNRTCSMPYA---KLEYKRSSNDLSNVSSSDGGR 976
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 631 ASSGTRITRLRPNDKSLQSKS--QOVAAAEHANKL-FKEGKEVLVN----- 677
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 GQMKPSIEISYSEDESKFCGYQYPRADLAHKITSANHMNDNDGLPPIVSLKYSDEQL 1036
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 678 SOGEISRLSTKKEVIMKGINNTYFKLQOEGKYRYHNOYSTNSFALNKHQHREDHK-RR 736
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1037 NSGRQSPSONEMRARPHIIEDEIKQSEQRQSR---NOSTT--YPVYTESTDKNHLPOR 1091
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 737 HLAHKFLTPAGFEKMGVSHGSKVLITSLRLTLTIOLENNITSSFPHPMAHBRANWIK 796
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1092 HFGQOECVSPYSRSGANGS-----ETNRVGS-----NHGINQNV 1125
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 797 AYOMCSKREFALALALECAVVPVWLPJWREFLGHTLRHRTSIREKEKVKKKKK 856
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1126 SQSLQED-----DYEDKPTNYSERY 1147
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 857 QEEETWQATWVKYTFPVKHOVWKOKGEYRYVYGGMSKITHVRFPLPGNTVY 916
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1148 SEED-----QHEEEDRPNTY-----SI 1164
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 917 NYRKSLEGTKNNMDENMDES-----DKRCKSRPKKIKIIEPDSEKDEV 959
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 1165 K N E E --- K R H D P I D I S L K Y A D D I P S S O K O S E F S K S S G O S S K T E H M S S S S N T S T 1220
Oy 960 K S D A K G A D Q N E M D I S K I T E K K D O V K E L L D S D S ----- D K P - C K E E P M E V 1005
Db 1221 P S M N K R --- Q N O L H P S S A O S R S G O P Q K A T C K V S S I N E T I O T Y C E D T P I C F S C S S L 1277
Oy 1006 D D D M K T E S H V N C O E S S Q ----- V D V V N S E G F H L R T ----- S Y K K K T S K U 1047
Db 1278 S S L S A E D I G C N O T Q E A D S A N T L O I A I E I K E I G T R S A E D P V S E V P A V S O H P R T K S S R L 1337
Oy 1048 D G L L E R I R I O F L E E K O R L E K I K E G I G I G I T S N S K N ----- L S S P V I - T K A E G 1101
Db 1338 Q C ----- S S L S E S A R K H A V E F S S G A K S P S K S G A Q T P K S P E H V O G T P L M F S R C T S V 1390
Oy 1102 C O S D S K R O E S P N A N D O P E D I L O C S O S D S V L R M S D P S H T T N K L Y P K D R V L D V S I R S 1161
Db 1391 S S L D S F E S H S I A S V S G E P ----- C S G M V S G I I S P S D L P D S G O T M P S R ----- 1435
Oy 1162 P E T K C P K O N S I E N D I E E K V S D L A S R G O E P T K S K T K G N D F I D S K L A S A D ----- 1211
Db 1436 S T P P P P O T A O T K R E ----- V P K N K A P T A E K R E S G ----- P K O A A V M A A V O R V O V L P 1483
Oy 1212 D I G T L I ----- C K N K ----- K P L I Q E S D T I V ----- S S K A L H S S V R K 1246
Db 1484 D A D T L H F T E S T P D G F S C S S L S A L S L D E P L O K D V E L R I M P P V O E N D N G M E T S E S O R K 1543
Oy 1247 S T N D R D A P L S R A M P E F K L G C D S E S N S T L E N S D T V S I O D S E E D M I O N S N E S I S E O F 1306
Db 1544 E S E N E Q ----- E K E A K T I D S E K D ----- L L D S D D D ----- 1570
Oy 1307 R R E D O V E L E P L K E L V S G E S T G C E D R L P V K G T P A N G K K S O O K L E B R P V N K S C S O I 1366
Db 1571 ----- D I E T L E --- E C - I I S A ----- M P T K S S R - K A K K P A O T A S K L P P P A R K P S O L 1612
Oy 1367 ----- K I K N T D K N N E R E S E K K G O R T S F O I N K O N K P I Y L K B E L K E I S E R V I S G A V 1423
Db 1613 P Y K K L P S O N R I Q O P K H V S F T P G ----- D D M P R V Y ----- C V ----- E C T P I N F S T 1653
Oy 1424 E F K V N N I N K I I P E N D I K S L T V K E S A I R P I N G D V I M E D F E N E R N S E T K S H L S S D A E G N 1483
Db 1654 A N S L S D L T I E S P N E L ----- A A G E V R G A O S G E F E K R D T I T T E G R --- S T D A O G G 1703
Oy 1484 Y N D S L E T L P S T K E S ----- D S T O T T P P S A C P E S 1512
Db 1704 K T S S V - T I P E L D N K A E E G D I L A E C I N S A M P K G S H K P P R V K K I M Q Y O G A S S S A P N K 1762
Oy 1513 N E V N O V E D M E I T S E Y K K Y T S ----- S P T S E E S N I S N D F 1548
Db 1763 N O L D G K K - K K P T S P V K P I P O N T E Y R T R V R K N A D S K N N L N A E R V F S D N K D S K O K N K N S 1820
Oy 1549 I D E N G L P I N K N E N V A G E ----- S K R K T V 1571
Db 1821 K P N D K L P N N E R V A G S F A F D S P H H T P I E G T P Y C F S R N D S L S L D P D D D V D L S E K --- 1878
Oy 1572 I T E V T M T S V A T E S K T V I K E V K O Q T V A S S T E N C A K S T V T ----- T T T T V T K 1621
Db 1879 - A E L R A K A E N K S E A K V T S H T E L T S N O S A N K T O A I A K O P I N R G O P K P I L O G S T P O G S 1937
Oy 1622 L S T P S T G S V D I S Y K E S K T V V T T V D S L T T G T I A T S M T V S K E Y S T R O K V I M K R S 1681
Db 1938 K D I P R G A A T D --- E K L O F A I E N T P V C S --- H N S S L S L S D I O E N N K N E N E P I K E T E 1991
Oy 1682 R P ----- K T R S G T A L P S Y R K ----- F V T K S T K S I F V L P N D L ----- K 1716
Db 1992 P P D S G E S K P A S G Y A R S F H V E D T P V C F S R N S S L S I D S E D D L O E C I S A M P K K 2051
Oy 1717 K L A R R G I R E V Y F Y N A K P ----- A L D I M P Y S P R P T G I T ----- W R Y R L 1758
Db 2052 K P S R L K D N E ----- K H S P R N W G I I G E D L F L D K I D R P D S E H L G S D S E N F D M K A I Q 2105
Oy 1759 Q T V K S L A G V S L M L R L I M A S L R M D D M A A K Y P P G G S G S T R E T S E T I T T T T I I R R D V G P Y G 1818

```

```

Db 2106 E G A N S I V ----- S S L H Q A A A A C L ----- S R O A S S D S D ----- S T L S L K S ----- G 2141
Oy 1819 I R R E Y C I R I K I C P I G V P E T P E K T P O R K G L R S S A L R P K R E T P K O T G P V I I E T W A E E E 1878
Db 2142 I ----- S I G S P F I L T P P O E K P F T S N K G P R --- I L K P E K S T ----- L E T K I E S E 2184
Oy 1879 ----- L E I M E I R A F E R V E K E K A O V E O A K R L E O Q R P ----- T V I A T S T 1919
Db 2185 S K G I N G K K Y K K S L I N G K V R N S E ----- I S Q M K O P L O A N N P S I R G R T M I H I P G 2235
Oy 1920 T S P T S T S T I S P A O K V M A P I S G V T T G K M L T I T K V S P A V T T V O O K N F O T R A T W Y 1979
Db 2236 V R N S S T S P S K K G P L K T P A K S P S E G O T A T T S P G A K P S V ----- 2278
Oy 1980 K O G O S N G V V O A K V L G I I P S S T G S O O T F T S F O R F A T V T R P N T S G S G G T S N S Q V I 2039
Db 2279 --- K S E L S V A R T S O I G ----- G S S K ----- A P R S G S R D S T P S ----- 2310
Oy 2040 T G P O I R P G M V I R P L O A S T L G K A I I R T P V M V O P G --- A P O Q V M O I I R G O P V S T A V S A 2095
Db 2311 ----- R P A O Q P L S R P I O S P G R N S ----- I S P G R N G I S P P N K L S Q L P R T S S P S T A --- 2354
Oy 2096 P N T V S T P O K S I T S ----- A T S N I Q S A S O P P R P O G O V K I T M A O L T O L T O G H 2146
Db 2355 - S T K S S G S G M S T S P G R O M S O O N L T K O F G L S K N A S S I P S E S A S --- K G I N O M N G N 2408
Oy 2147 G C N O G L T V V I O G O G O T T G O L Q ----- L I P O G V T V L P G G O O L M O A M P N G T V O R F L E T 2199
Db 2409 G A N K K V E L S M S T K S G S S D S E R P V L R O S T F I K E A P S P T L R R K L E S S A S ----- F E 2463
Oy 2200 P L A T T A T T A S T T T V S T A A C T G E O R O S K L S P ----- O M O V H O D ----- K T L P P A O S S 2248
Db 2464 S L S P S R P A S P T S O A Q T E V ----- L S P L P D M S I S T H S V O A G M W R K L P N L S P 2513
Oy 2249 S V G P A K A Q O T A P S A R P Q P O P O S P A P E V O T O P E V O T Q T V S H V P S E A O P T A O S S 2308
Db 2514 T I E Y N D G R P A K R H D I A R S H S E S P S R L P I N R S G T W K R E --- H S H S S L P L P V S T W R R T G S S 2570
Oy 2309 K P O V A A S O P O S N V O G O S P V R V O S P S O T R I R P S T P S O L S - P G O O S V O Q T T S O P I R O P H 2367
Db 2571 S S T L S A S S E S E K A K S E D E K H V M S I S G T --- K O S K E M O V A K G I W R I K E N E F S P ----- 2622
Oy 2368 T S L O I P S G O P O S O P O V O S S T O T L S S G O T 2396
Db 2623 ----- T N S T S O T V S S G A T 2635

RESULT 15
US-09-003-687A-7
; Sequence 7, Application US/09003687A
; Patent No. 5998600
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/003,687A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,355
FILING DATE: 20-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UT
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5998600e
US-09-003-687A-7

Query Match 2.4%; Score 362.5; DB 2; Length 2973;
Best Local Similarity 17.3%; Pred. No. 8,6e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

QY 373 WEAELEKILEEMEEIHRHDDITEDLTNKRAGSKSFLLAANEELLESIRAKGIDIDNKA 432
DB 699 WDGAVSMKLNLIHSRHKM-----AMGS-----AALLNLNANRPKAKYKA-NIM 743
QY 433 SPEET-----EKDN-ETENDSKDAEKNEEFEDQSLKSDSDKTDDPEQKSEEP 485
DB 744 SPSSSLPLHLVRKOKALEHLDQHLSSETFDNIDNLSPKASHSKQRHKOSLYGDYVFD 803
QY 486 EYGDCKNSVANIGD-----NTT-----NATSEETSPSEGRSPVC-- 521
DB 804 NRHDDKRSNFTNGMTVLSPYLTNTVLSPSSSSSGSLDSSRSEKRSLEEREGIGLGN 863
QY 522 --LSETPDSSNMAKKVASELPQ--DYPEPN--KTCESSNTSATPTS-----QNLN 570
DB 864 HPTETPGTSSKRGLOISTIAOIAKAMEVSAIHTSOEDRSSGSTTELCVTDENALR 923
QY 571 SNSSSELNQSOSAKAADPENGEREHTPVSIQEIYODFTSEKSTGELSESPAGK 630
DB 924 RSSAAHTSHNTYFTKS---ENSNTCSMPYA---KLEYKRSSNDLSNVSSSDGYGR 976
QY 631 ASGSTIILIRLNPDSKLSQKS--QQVAAAHEANKL-FKSGEVLVYN----- 677
DB 977 GQMKPSIESYSEDDSEKFCGYQYRADLAKIHSAHMDNDNGELDTPINYSILKYSDEOL 1036
QY 678 SOGEIIRLSTKREKVINNNYFKLGQSKYVYVHNOYSTSEFALNKHQHRDHK-RR 736
DB 1037 NSGRQSPQONERAKPKHIIEDIKOSEORQR--NOSTT--YVYTTSTDKHLKFGP 1091
QY 737 HLAHKECLTPAGEFKWNGSVHSGKVLITSLRLITOLENNIPSPSFPHNMAASHRANWK 796
DB 1092 HFGQGCVCSPYRSGANGS-----ETNRRGS-----NHGYNQV 1125
QY 797 AVQMCSEPEEFALALILECAVKPVYMLPIWREFLGHTRLHMTSIEREKEKVKKEK 856
DB 1126 SOSLCQED-----DYEDDKPTNYSERY 1147
QY 857 OEEEFMQATWYKTFYVYKHOVWVKQGEERYVYGYGSMISKTHVYFVPRKLPENTVY 916
DB 1148 SEEE-----OHEEERPTNY-----SI 1164
QY 917 NYRKSLEGTAKNNMDENMDES-----DKRCSRSPKTIKIPDSEKDEY 959
DB 1165 KNEE-----KRHYDOPIDISLKYAIDIPSSOKOSFSPKSSSGQSSKTEHMSSENTST 1220

QY 960 KGSDAKAGADONEMDISKITEKKODVYKELLSDS-----DKP-CKEEPMY 1005
DB 1221 PSSNAKR---QNDLHSSASRSGQFOKATCAVSSINDETIOTYCEVDPIFCSSSL 1277
QY 1006 DDDMKTESHVNCQSSO-----VDVNVSEGFHLRT-----SYKRTKSKU 1047
DB 1278 SLSLSAEDETICQNOTQEADSANLQIAELIKERIGTSRADDPYEVPAVSQHRTRKSR 1337
QY 1048 DGLLERIKOFTLEEKQORLEKILBEGIKIGITSNSSKN-----LSEPPY-TRAKES 1101
DB 1338 QG-----SLSSESARHKAVEFSSGAKSPSKSGAQTSPPEHYVQETPLMFHSTVS 1390
QY 1102 QSDSMKROEQSPANNQDPEDLIOGCSQSDSVLRRASDSHTNKLXPDRVLDVDSIRS 1161
DB 1391 SLSDFESRSIASSVQSEF-----CSGWSGIIISPDLPDSFGQTMPRSR----- 1435
QY 1162 PETCKPKNSIENDIEKVSDLASRQOEPTKSKTKCNDFIDSKLASAD----- 1211
DB 1436 SKTPPPPTQATKRE-----VPKNKAPTAEKRESG-----PKOAAVMAAVORYVLP 1483
QY 1212 DIGTL-----CKNK-----KPLIOESDITV-----SSKSALHSSVPK 1246
DB 1484 DADTLHPTATESPPDQSCSSSLASLDBPFIQKVDELRIKMPVOENNGNETSEQPK 1543
QY 1247 STNDRDATPLSRAMDEFGKIGCDSESNSTLSSSDTVSIOSSSEEDMIYONSSESIOF 1306
DB 1544 ESNENQ-----EKAETKIDSKD-----LLDSDDD----- 1570
QY 1307 KREQOVLEVLEPKELVSGESTGNCEDRLPVKGTANGKKBQCKLEERYNKCSDOI 1366
DB 1571 ----DIETLE-EC-IISA-----MPTKSR-KAKKAPQATSKLPPPARKPSOL 1612
QY 1367 ---KLNTDKKNENRESKKGQRTSTFOINKDKPKITYLKGELKTESRNVSGV 1423
DB 1613 PYYKLPLSQNRLOPKHVSFTG-----DMPRVY---CV-----EGTPINEST 1653
QY 1424 EPRVNNINKIIPENDIKSLTVKESAIRPINDVIMEDENERSSEFKSHLLSSDAEGN 1483
DB 1654 ATSLSDLTIESPNEI-----AAGEVYRGAGQGEFEKRDITPTGR--STDAQCG 1703
QY 1484 YRDSLETLPTSKES-----DSQTTTPSACPS 1512
DB 1704 KISSV-TIPELDNKAEBGDIABECLINSAMPKSKHPPRVKIMQVOASASSAPAK 1762
QY 1513 NSYQVEDMEIETSEYKAKYTS-----SPITSEESNLNDF 1548
DB 1763 NQDQKK--KKPTSPVKPIPNQTEYRTVRKNAKSKNNINAEVFPDNNKSKQNNKNS 1820
QY 1549 IDENGULPINKENYNGE-----SKRTV 1571
DB 1821 KQFNDKLPNNMEDVRGSAFADSPHHTPIEGTPYCFSRNDSLSLDFDDDVLSREK-- 1878
QY 1572 IREVTMTSTVATESKTYVKEKGDQYVASSPENCASKSTV-----TTTTTVK 1621
DB 1879 -AELRAKENKESAEKAVYHTLETLTSMQSANQOIAIAKOPINRGQPKPILOKOSTFPOSS 1937
QY 1622 LSTPTSGASVDIISVKEQSKTYVTTLTSLTTGGTILVTSMTVSKESYSTRDKVKLMKS 1681
DB 1938 KOIPDRGAATD--EKLOMPAIENTPVCS---HNSLSLSLDIOENNNKKEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FVTKSTKSTFVLPNDL-----K 1716
DB 1992 PPSQSGEPKPOASGVAPKSFHVEDTPVCFSRNSSLSSISDEDDLLOECISSAMPKK 2051
QY 1717 KLAARKGIREVYFNNNAK-----ALDIMPYSPRPPTGIT-----WRRL 1758
DB 2052 KPSRLKGNE-----KSPRMNGGILGEDTLIDKLDIQRPSEBGLSDSSENFPMKAIQ 2105
QY 1759 QYVKSILAGVSLMLRLMLASLRMDDMAKVPPEGGSTRTESETETTTTIIKRRDVPYG 1818
DB 2106 EGANSTIV-----SSLHAAAAACL-----SKAASSSD-----SILSKS-----G 2141
QY 1819 IRPEYCIIRKIICPIGVETPKETPPQQRKGLRSSALRPKRPTPROTGPIVLIETWAAEE 1878

```
Db 2142 I-----SLGSPFHLPPOEEKPFTSNKGPR--ILKPGEKST-----LETKKIESE 2184
QY 1879 /-----LELMEIRAFARVEKEKAQAVEOAKKRL EOQKP-----TVIATSP 1919
Db 2185 SKGKKGKVKVYKSLTGKVRNSNE-----ISGQMKOPLQANMPISIRGRTMHIIPG 2235
QY 1920 TSPSTSTTISBAQKVMAPIISGAVTTGTKNVLTTKVSPATVTPQONKNFHOTFATWY 1979
Db 2236 VRNSSSTSPVSKGPPPLKTPAKSPSEQOTATTSRGAKPsv-----2278
QY 1980 KQGSNSGVVOQKVLGITPSSTGTSQOTFSPQRTATVTIRPNTSGSGGTTNSQVY 2039
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APSRSGSRDSTPS--- 2310
QY 2040 TSPQIRPGMTVIRTPLOQSTLGKAILRTPVMVPG---APOQVMTQIRGQPVSTAVSA 2095
Db 2311 -----RPAGQPLSRPIQSGRNS-----ISPGNGISIPNKLSQLPRTSSPSTA--- 2354
QY 2096 PNTVSTPQOKSLTS-----ATSTSNIOSSASQPPRPQOGVYKLTWAOULTOLTOGH 2146
Db 2355 -STKSSGSGKMSYTSPIGRMSQONLTKOTGLSKNASSIPRESAS-----KGLNQMNNGN 2408
QY 2147 GGNQGLTVVIOGQGTGQLQ-----LIPGVTVLPBGQOLMGAAMPNGTVQRFLET 2199
Db 2409 GANKKVELSRMSTKSSGESDSRSERPVLVROSTFIKEAPSTLRKLEESAS-----FE 2463
QY 2200 PLATTATTAATTTTIVSTTAAGTGEOROSKLSR-----OMOVHOD-----KTLPPAOS 2248
Db 2464 SLSPSRPASPTRSQAKQTFV-----LSPSLPDMSLSTHSSVQAGGWRKLPNLSR 2513
QY 2249 SVGPAKAQOPQTAQPSARPOPOQPOSPAQPEVOTQPEVOTQTVSSHVPSEAOPTHAOS 2308
Db 2514 TIEYNDGRPAKRDHDIARSHSESPSLPILNRSGTWKRE---HSKHSSSLPRVSTWRTGSS 2570
QY 2309 KPQVAAQSQPQSNVQGSFVRVQSPSQTRIRPSTPSQLS--PGQSQVQTTTSQPIPIQPH 2367
Db 2571 SSTLSASSSESEKASEDEKHVNSISGT--KQSKENQVSAKGTWRKIKENESP----- 2622
QY 2368 TSLQIPSOQPOPOPOVOSSTQTLSSGQT 2396
Db 2623 -----TNSTSQTVSSGAT 2635
```

Search completed: November 20, 2002, 16:32:15
Job time : 105.976 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:05 ; Search time 74.106 Seconds
(without alignments)
5227.100 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MWSEEEEDGDAEETDSE.....KIKGFKASRSHNNKIQSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the prediction being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14971	100.0	2907	21	AAV57452 Human transcriptio
2	14170	94.6	2781	21	AAV57453 Human transcriptio
3	2695.5	18.0	557	22	ABB94078 Human protein sequ
4	1488.5	9.9	976	22	ABB68735 Drosophila melanog
5	1385	9.3	328	21	AAV54364 Human pancreatic c
6	982.5	6.6	238	20	AAV04323 Human secreted pro
7	981	6.6	237	22	AAU19446 Human diagnostic a
8	789	5.3	249	23	AAU78845 SCAN/KRAB protein
9	781	5.2	246	22	AAU31944 Novel human secret
10	661.5	4.4	997	22	ABB62887 Drosophila melanog

11	621.5	4.2	2768	22	ABB68397 Drosophila melanog
12	613.5	4.1	5533	22	ABB65772 Drosophila melanog
13	613.5	4.1	5560	22	ABB71160 Drosophila melanog
14	567	3.8	6815	22	ABB66811 Drosophila melanog
15	562.5	3.7	3257	22	ABB67502 Drosophila melanog
16	561	3.7	3080	22	ABB64877 Drosophila melanog
17	551.5	3.7	2897	22	ABB58514 Drosophila melanog
18	503.5	3.4	4498	22	ABB58514 Drosophila melanog
19	491	3.3	2137	23	ABP39618 Staphylococcus epi
20	474	3.2	3726	22	ABB63947 Drosophila melanog
21	462	3.1	2586	22	ABB66878 Drosophila melanog
22	458.5	3.1	3266	21	ABA42491 Human OREF ORF2255
23	449.5	3.0	3111	22	ABB60327 Drosophila melanog
24	444	3.0	2977	22	ABB69480 Drosophila melanog
25	443	3.0	2344	22	AAU37120 Staphylococcus aur
26	441.5	2.9	3696	23	ABP40235 Staphylococcus epi
27	439.5	2.9	1878	22	AAV40239 Human polypeptide
28	439.5	2.9	3256	21	AAV50976 Human cell cycle p
29	435	2.9	1727	22	ABB95554 Human protein sequ
30	433	2.9	1795	22	ABB69806 Drosophila melanog
31	432	2.9	2665	22	ABB28314 Human peptide #965
32	432	2.9	2665	22	ABB33490 Human peptide #965
33	432	2.9	2665	22	ABB18950 Protein #949 encod
34	432	2.9	2665	22	AAV54270 Human brain expres
35	432	2.9	2665	22	AAV66665 Human bone marrow
36	432	2.9	2665	22	AAV14533 Peptide #967 encod
37	432	2.9	2665	22	AAV26950 Peptide #987 encod
38	432	2.9	2665	22	AAV02259 Peptide #941 encod
39	432	2.9	2665	23	ABB36319 Human peptide enco
40	429.5	2.9	3201	22	ABB62899 Drosophila melanog
41	427	2.9	2951	22	ABB60291 Drosophila melanog
42	423	2.8	2285	22	ABB63057 Drosophila melanog
43	417.5	2.8	1878	19	AAW81170 Human BAZ2-alpha P
44	417	2.8	2441	21	AAV18161 Plasmodium falcipa
45	414.5	2.8	3169	22	ABB64569 Drosophila melanog

ALIGNMENTS

XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

PS Claim 1; Page 48-68; 154bp; Japanese.

CC The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulatory mechanism. It binds
CC to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.

XX Sequence 2907 AA;

Query Match 100.0%; Score 14971; DB 21; Length 2907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSEEEEDDGAETQDSDEDEDEDDDDSDYPREMEDDDDDASYCRESSFRSHST 60
DB 1 MYSEEEEDDGAETQDSDEDEDEDDDDSDYPREMEDDDDDASYCRESSFRSHST 60
QY 61 YSSTPGRRKPRVHRPSPILIEKDIPLLEPKSSEDLAVPEHIMNVAIYEVRNFGTV 120
DB 61 YSSTPGRRKPRVHRPSPILIEKDIPLLEPKSSEDLAVPEHIMNVAIYEVRNFGTV 120
QY 121 LRLSPFRREDCALVSOEQTLMABHVVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
DB 121 LRLSPFRREDCALVSOEQTLMABHVVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
QY 181 FIDGTMPEVLRVYCESDKEVHNVLPOEADYPGVPENKIKYLOFLVDOFLTNTARE 240
DB 181 FIDGTMPEVLRVYCESDKEVHNVLPOEADYPGVPENKIKYLOFLVDOFLTNTARE 240
QY 241 ELMSEGVLYQYDHCRCVCHKLGLDLCETCSAVHLECVKPRLEBVPEDEMOCEYVAHKV 300
DB 241 ELMSEGVLYQYDHCRCVCHKLGLDLCETCSAVHLECVKPRLEBVPEDEMOCEYVAHKV 300
QY 301 PGVYDCAEALOKNKPYIRHEPIGYDRSRKRYFLNRLLIIEEDTENNEKKIWTYSTKVQ 360
DB 301 PGVYDCAEALOKNKPYIRHEPIGYDRSRKRYFLNRLLIIEEDTENNEKKIWTYSTKVQ 360
QY 361 LAELIDCLDKYMAELCKILEEMREIHRHMDITTEDLTNKRGSNKSFLAANEELIES 420
DB 361 LAELIDCLDKYMAELCKILEEMREIHRHMDITTEDLTNKRGSNKSFLAANEELIES 420
QY 421 IRAKGDIDNVKSPETIEKDKNETENDSKDAEKNREFEEDOSLEKSDDKTPDDPEQK 480
DB 421 IRAKGDIDNVKSPETIEKDKNETENDSKDAEKNREFEEDOSLEKSDDKTPDDPEQK 480
QY 481 SEEPTEVGDKGNSVNLGDTTMTASEETSPSRGRSPVGLSTPPSSNMAEKKVASEL 540
DB 481 SEEPTEVGDKGNSVNLGDTTMTASEETSPSRGRSPVGLSTPPSSNMAEKKVASEL 540
QY 541 PDVPEEPNKTCESSNTSATTTSTIOPNLSENSSSSELNSSSESASAKAADPENGERESH 600
DB 541 PDVPEEPNKTCESSNTSATTTSTIOPNLSENSSSSELNSSSESASAKAADPENGERESH 600
QY 601 PVSIOEELVGDFTSEKSTGELSESFGAGKAGSSTRITTRLRNPDSLSOLKSOQVAAA 660
DB 601 PVSIOEELVGDFTSEKSTGELSESFGAGKAGSSTRITTRLRNPDSLSOLKSOQVAAA 660
QY 661 HEARKLEKEGEVLYVNSOGELISRLSTKKEVYKNGTNNYFKLGOEKKYVYVHYOYSTN 720
DB 661 HEARKLEKEGEVLYVNSOGELISRLSTKKEVYKNGTNNYFKLGOEKKYVYVHYOYSTN 720
QY 721 FALNRHGHREDHDKRRLAHKFCITPAGEFKWNSVHGSVLTISTLRLTITOLENNIPS 780
DB 721 FALNRHGHREDHDKRRLAHKFCITPAGEFKWNSVHGSVLTISTLRLTITOLENNIPS 780
QY 781 SFPHPNASIRANVKAIVQMSKPREFALALALIECAVKFVVMPLVIREFLGHTRLHMT 840
DB 781 SFPHPNASIRANVKAIVQMSKPREFALALALIECAVKFVVMPLVIREFLGHTRLHMT 840
QY 841 SIEREEKVKKKEKKKEEETMOATWVKYTFPVKHOVWKOGKEEVRVYGYGWSMTSK 900

DB 841 SIEREEKVKKKEKKKEEETMOATWVKYTFPVKHOVWKOGKEEVRVYGYGWSMTSK 900
QY 901 THVRFVPKPLRGNTNVYNRKSLLECTKNMNMDESDKRKRSRSPKRIKIPDESEKDEVK 960
DB 901 THVRFVPKPLRGNTNVYNRKSLLECTKNMNMDESDKRKRSRSPKRIKIPDESEKDEVK 960
QY 961 GSDAKAGADQNMEDISKTEKKDQDVKEELDSDSDKPCKEPMPVDDMTESIVNQES 1020
DB 961 GSDAKAGADQNMEDISKTEKKDQDVKEELDSDSDKPCKEPMPVDDMTESIVNQES 1020
QY 1021 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKORLEKILEGGIKIGK 1080
DB 1021 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKORLEKILEGGIKIGK 1080
QY 1081 TSTNSKNLSSEPVITAKECQSDSMRQESPANNDQPEDLIQGCQSQSDSVLRMSDP 1140
DB 1081 TSTNSKNLSSEPVITAKECQSDSMRQESPANNDQPEDLIQGCQSQSDSVLRMSDP 1140
QY 1141 SHTTNKLYPKDRVLDVYSIRSPETKCPKONSIENTIEEKYSDLASRQEPYTKSKTKGND 1200
DB 1141 SHTTNKLYPKDRVLDVYSIRSPETKCPKONSIENTIEEKYSDLASRQEPYTKSKTKGND 1200
QY 1201 FIDSKLASADDTLTLCKNNKPLIQEESDTYSSSKSAHSSVPKSTNDPAPLSRAM 1260
DB 1201 FIDSKLASADDTLTLCKNNKPLIQEESDTYSSSKSAHSSVPKSTNDPAPLSRAM 1260
QY 1261 DFEKGLGDSRSTLNSDPTVSIQDSSEDMIVQNSNISISQFTRQDQVEVLEPLK 1320
DB 1261 DFEKGLGDSRSTLNSDPTVSIQDSSEDMIVQNSNISISQFTRQDQVEVLEPLK 1320
QY 1321 CELVSGESTGNCEDRLPVKGTETANGKRPQOQKLEEBPVKCSQDIKLTNTTKKNNE 1380
DB 1321 CELVSGESTGNCEDRLPVKGTETANGKRPQOQKLEEBPVKCSQDIKLTNTTKKNNE 1380
QY 1381 ESEKKGRTSTFOJNGDNKPKIYLKGECLKEISESVYSGVNEPKYNNINKIIPENDIK 1440
DB 1381 ESEKKGRTSTFOJNGDNKPKIYLKGECLKEISESVYSGVNEPKYNNINKIIPENDIK 1440
QY 1441 SLVYKESAIRPFINGVIMEDFNERNSSEFKSHLSSDDEGNVRSLEFLPTKESDST 1500
DB 1441 SLVYKESAIRPFINGVIMEDFNERNSSEFKSHLSSDDEGNVRSLEFLPTKESDST 1500
QY 1501 QTTTPSASCEPSNVNOVEDMEIETSEVKYVTSPTISEESNLNDPIDENGILPKNBE 1560
DB 1501 QTTTPSASCEPSNVNOVEDMEIETSEVKYVTSPTISEESNLNDPIDENGILPKNBE 1560
QY 1561 NVNGESKRKTIVTEVTMTSTVATESTVIKVEKGQOTVYVSSBENCAKSTVTTTYYV 1620
DB 1561 NVNGESKRKTIVTEVTMTSTVATESTVIKVEKGQOTVYVSSBENCAKSTVTTTYYV 1620
QY 1621 KLSPTSGVSDIISVKEQSKTVTTVTJDSLTGTLVTSMVSKESYSTRQVKLMKF 1680
DB 1621 KLSPTSGVSDIISVKEQSKTVTTVTJDSLTGTLVTSMVSKESYSTRQVKLMKF 1680
QY 1681 SRPKTRSGTALDPSYKFEVYKSTKKSIFVLPNDDLKLRKGGIREVYFYNNAKPAIDI 1740
DB 1681 SRPKTRSGTALDPSYKFEVYKSTKKSIFVLPNDDLKLRKGGIREVYFYNNAKPAIDI 1740
QY 1741 WPYSPRPFTGIMRWRYLOVYKSLAGVSLMLRLIMASLRMDMAAVPPEGSGSTRTESE 1800
DB 1741 WPYSPRPFTGIMRWRYLOVYKSLAGVSLMLRLIMASLRMDMAAVPPEGSGSTRTESE 1800
QY 1801 TELTTEELIKRQVGPAGIFEEYCIRKIIICPIGVETPEKTPPQPKGLRSSALRKRBE 1860
DB 1801 TELTTEELIKRQVGPAGIFEEYCIRKIIICPIGVETPEKTPPQPKGLRSSALRKRBE 1860
QY 1861 TPQOTGVLIETVVAEELELMEIRAFEREREKAQAEQAKRLEQOKPTVIATSTT 1920
DB 1861 TPQOTGVLIETVVAEELELMEIRAFEREREKAQAEQAKRLEQOKPTVIATSTT 1920
QY 1921 SPTSSTSTISPAQVWVAPISSVYTGKMWLTITVGSAAVTPPOAKNFHOTFTATWVK 1980

Db 1921 SPTSSTTSTISPAQKVMAPISGSVTTGTAKMVLTTVGSFPAIVTFQONKNHOTEFTATWAK 1980
QY 1981 OGQSNQSVVVOVQOKVLGIIPSSSTGTSTQOTFTSPQPTATVTITRPNSTSGSGCTTNSQVIT 2040
Db 1981 QGQSNQSVVVOVQOKVLGIIPSSSTGTSTQOTFTSPQPTATVTITRPNSTSGSGCTTNSQVIT 2040
QY 2041 GPOIRGKMTVIRTPLOOSTLGAIIITPVWVOPGARQOVNTQIIRQOPVSTAVSAPNTVS 2100
Db 2041 GPOIRGKMTVIRTPLOOSTLGAIIITPVWVOPGARQOVNTQIIRQOPVSTAVSAPNTVS 2100
QY 2101 SPPGOKSLTSASTNSINQSSASQPPRPOGQVNLTAOLTLQTLQGHGNGGLTVNVOGQ 2160
Db 2101 SPPGOKSLTSASTNSINQSSASQPPRPOGQVNLTAOLTLQTLQGHGNGGLTVNVOGQ 2160
QY 2161 QTTGQQLIPQGVTVLPBGQOQLMAAMPNGTVQRFLETPRLATTATTAATTTTSTTAA 2220
Db 2161 QTTGQQLIPQGVTVLPBGQOQLMAAMPNGTVQRFLETPRLATTATTAATTTTSTTAA 2220
QY 2221 GTGEOQROSKLSPQMOVHODKTLTPRAOSSSVGPAKAQPTQPSARPOTQPSAPQPEV 2280
Db 2221 GTGEOQROSKLSPQMOVHODKTLTPRAOSSSVGPAKAQPTQPSARPOTQPSAPQPEV 2280
QY 2281 QTPQEVQOTQTVSSHVSEAPQTHAASSKQVAAQSQPQSNVQSGSPVRYQSPQTRIRP 2340
Db 2281 QTPQEVQOTQTVSSHVSEAPQTHAASSKQVAAQSQPQSNVQSGSPVRYQSPQTRIRP 2340
QY 2341 STPSQSLSPGQSQSOVQTTTSPQIPQPTSLQIPSGQPOQSQPOVQOSTQTLSSGOTLNV 2400
Db 2341 STPSQSLSPGQSQSOVQTTTSPQIPQPTSLQIPSGQPOQSQPOVQOSTQTLSSGOTLNV 2400
QY 2401 SVSSPSPRPOQLIQPPQVIAVBPOLQOQVQVLSQIQSQVVAQIQAOQSGVPOQIKLQLP 2460
Db 2401 SVSSPSPRPOQLIQPPQVIAVBPOLQOQVQVLSQIQSQVVAQIQAOQSGVPOQIKLQLP 2460
QY 2461 QIQOQSSAVQTHQIQNVVTVQAASVQEOLOQVQOOLRQOQKKQOQIETINNTSPSKLITV 2520
Db 2461 QIQOQSSAVQTHQIQNVVTVQAASVQEOLOQVQOOLRQOQKKQOQIETINNTSPSKLITV 2520
QY 2521 EITQOVVKNHNAVIEHLKQKSMTPAEREENQMIYCNQVMKYIIDKIDKEEKOAAKRR 2580
Db 2521 EITQOVVKNHNAVIEHLKQKSMTPAEREENQMIYCNQVMKYIIDKIDKEEKOAAKRR 2580
QY 2581 KRESEYEQKRSKONATKLSALLFKHKEQLRAELIKKRALDKDLQIEVQBELKRDLIK 2640
Db 2581 KRESEYEQKRSKONATKLSALLFKHKEQLRAELIKKRALDKDLQIEVQBELKRDLIK 2640
QY 2641 EKQMLQATATAAARCPRTVLPAPRAPRPPRPPRGVQNHGGLSTPLPVAASQKRR 2700
Db 2641 EKQMLQATATAAARCPRTVLPAPRAPRPPRPPRGVQNHGGLSTPLPVAASQKRR 2700
QY 2701 EEEKDSSSKSKKKMISTSKETKMDTKLYCICKTPYDESKFYIGCDRCOMWTHGRCVGI 2760
Db 2701 EEEKDSSSKSKKKMISTSKETKMDTKLYCICKTPYDESKFYIGCDRCOMWTHGRCVGI 2760
QY 2761 LQSEAEILIDEYVQPOQOSTEDAMTVLPLTEKDYEGIKRVLSIQAHKMAWPLEPVDN 2820
Db 2761 LQSEAEILIDEYVQPOQOSTEDAMTVLPLTEKDYEGIKRVLSIQAHKMAWPLEPVDN 2820
QY 2821 DADYDYGVIKEPDLATMEERQVRRYEKLTEVADMTKIFDNCRYNPSDSFYOCAEV 2880
Db 2821 DADYDYGVIKEPDLATMEERQVRRYEKLTEVADMTKIFDNCRYNPSDSFYOCAEV 2880
QY 2881 LESFVQKLKGFKASRSHNNKLOSTAS 2907
Db 2881 LESFVQKLKGFKASRSHNNKLOSTAS 2907

RESULT 2
AA57453
ID AA57453 standard; Protein; 2781 AA.
AC XX
XX AA57453;
XX

DT 22-FEB-2000 (first entry)
XX Human transcriptional regulatory factor SEQ ID NO:10.
XX Human transcriptional regulatory factor: TCOA1, BLAST detection:
KW bromo-domain; cell proliferation; cancer.
OS Homo sapiens.
PN M09957143-A1.
XX 11-NOV-1999.
PD 30-APR-1999; 99MO-JP02340.
PF 30-APR-1998; 98JP-0137631.
PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA Jones MH;
PI WPI: 2000-052940/04.
DR N-PSDB; AA239033.
PT Transcriptional regulatory factor containing a bromo domain and gene
PT TCOA1 encoding it -
PS Claim 1; Page 139-151; 154pp; Japanese.
XX The present sequence represents a human transcriptional regulatory factor
XX containing a bromo domain. The factor interacts with proteins involved
XX in the chromatin-mediated transcription regulation mechanism. It binds
XX to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening
XX compounds binding to it and acting as agonists or antagonists, which
XX are potentially useful for the treatment and prevention of cancer and
XX other cell proliferation disorders.
SQ Sequence 2781 AA:
Query Match 94.6%; Score 14170; DB 21; Length 2781;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
QY 1 MYSEEEEDDGAEEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60
Db 1 MYSEEEEDDGAEEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60
QY 61 YSSTPGRRKPRVHRPNSPILEEKDIPLEFPKSSDLMVFNHIMVIAIYEVLRNFTV 120
Db 61 YSSTPGRRKPRVHRPNSPILEEKDIPLEFPKSSDLMVFNHIMVIAIYEVLRNFTV 120
QY 121 LRLSPREFDFCAALVSOECSTLMAEKNHVVLLKAVLREEDTSWTFGPADLKQSVNSTLY 180
Db 121 LRLSPREFDFCAALVSOECSTLMAEKNHVVLLKAVLREEDTSWTFGPADLKQSVNSTLY 180
QY 181 FIDGMTPREVLRVYCESDKEYNHVLRYQEAEDYRQVFNKIKVLOFLVQDLFTTNIAE 240
Db 181 FIDGMTPREVLRVYCESDKEYNHVLRYQEAEDYRQVFNKIKVLOFLVQDLFTTNIAE 240
QY 241 ELMSEGVIOYDDHCRCVHKLGDLILCCETCSAVYHLECVKPRLEVEPDEMQCEVCAHKV 300
Db 241 ELMSEGVIOYDDHCRCVHKLGDLILCCETCSAVYHLECVKPRLEVEPDEMQCEVCAHKV 300
QY 301 PGVTDCVAETQKMKRYIRHEPTGYDSSRRKYWFLNRLLIEEDTENENKKTWYSTKVO 360
Db 301 PGVTDCVAETQKMKRYIRHEPTGYDSSRRKYWFLNRLLIEEDTENENKKTWYSTKVO 360
QY 361 LAILDCLDKDYWEALCKLLEEMREEIHRHMDITEDLTNKAQSKSFLAANEELIS 420
Db 361 LAILDCLDKDYWEALCKLLEEMREEIHRHMDITEDLTNKAQSKSFLAANEELIS 420
QY 421 IRAKKDIDNVKSPETEKKNETENDSKDAEKNREEPDQSLKQSDKTPDDPEQK 480

421 IRAKGDIDNVKSPBETEKDNKNETENDSKDAEKNREFEQSLKSDDKTPDDPEQK 480
481 SEEPTEVGDKNSVSNLGDNTNATSEETSPSGRSPVGLSETPOSSNAEKKVASEL 540
481 SE----- 482
541 PODYPEENKTCESNNTSATTTSIOPLNENSSSELNSQSESAKADDPENGERSHT 600
483 ----- 482
601 PVSIOEIVGDFTSKSTGELSESPGAGKAGSGSTRITRLRNPDSKLSQKSOQVAAA 660
483 ----- 534
661 HEANKLFEGKEVLYVNSQGEISRLSTKEVIMKGINNYKLGQEGRYRYHNOYSTNS 720
535 HEAKKLFEGKEVLYVNSQGEISRLSTKEVIMKGINNYKLGQEGRYRYHNOYSTNS 594
721 FALKHQRHEDHDKRRHLAHFCLTPAGEFKWNGSVHSGKVLITSTLRLLTTOLENNIPS 780
595 FALKHQRHEDHDKRRHLAHFCLTPAGEFKWNGSVHSGKVLITSTLRLLTTOLENNIPS 654
781 SFPHPNASHRANMKAQOMCSKPREFALALALECAVKPVVMDPIWREFLGHTRLAHMT 840
655 SFLHPNASHRANMKAQOMCSKPREFALALALECAVKPVVMDPIWREFLGHTRLAHMT 714
841 SIERREKVKKKKEKQEEETMOATVVKYTFPVKHQVWKQKQBEVYVGTGGMSWISK 900
715 SIERREKVKKKKEKQEEETMOATVVKYTFPVKHQVWKQKQBEVYVGTGGMSWISK 774
901 THYVRFPVKLGNTNNVATKSLBGTKNMDEMDESDKRCSSRSFKKTIKIEPDSEKDEVK 960
775 THYVRFPVKLGNTNNVATKSLBGTKNMDEMDESDKRCSSRSFKKTIKIEPDSEKDEVK 834
961 GSDAKAGQOMEMDISKTTEKKDQVLELDSDPCKEPEMVEDDMKTESHVNOES 1020
835 GSDAKAGQOMEMDISKTTEKKDQVLELDSDPCKEPEMVEDDMKTESHVNOES 894
1021 SQVDVNVSEGFHLRTSYKKKTSSKLDGLERRIKOFTLEKORLEKIKLEGGIKGIGK 1080
895 SQVDVNVSEGFHLRTSYKKKTSSKLDGLERRIKOFTLEKORLEKIKLEGGIKGIGK 954
1081 TSTNSSKLSLSPYITRAKECCQSDSMQEOSSPNANNDQPEDLIQGCQSQSSSVLRMSDP 1140
955 TSTNSSKLSLSPYITRAKECCQSDSMQEOSSPNANNDQPEDLIQGCQSQSSSVLRMSDP 1014
1141 SHTNNKLYPKRVLDDVYSIRPEPKCPKONSIENDIEEKVSDLASRGOEPTKSTKGMDP 1200
1015 SHTNNKLYPKRVLDDVYSIRPEPKCPKONSIENDIEEKVSDLASRGOEPTKSTKGMDP 1074
1201 FIDDSKLASADDIGTLICKNKPLIOEESDPIVSSSKSALHSSVPKSTNDRDAPLSRAM 1260
1075 FIDDSKLASADDIGTLICKNKPLIOEESDPIVSSSKSALHSSVPKSTNDRDAPLSRAM 1134
1261 DFEKRLGDSNSTLEKNSSTVSIQDSSEEDMAYONSNEISIQOFRREDDEVLEPLK 1320
1135 DFEKRLGDSNSTLEKNSSTVSIQDSSEEDMAYONSNEISIQOFRREDDEVLEPLK 1194
1321 CELVSGESTGCEBRLPVKGTGEANGKPKSOOKLLEBRPVNKCSDQIKLAKTTDKKNENR 1380
1195 CELVSGESTGCEBRLPVKGTGEANGKPKSOOKLLEBRPVNKCSDQIKLAKTTDKKNENR 1254
1381 ESEKRGORTSTFOJNGKDNKPKIYLKGECLKEISSESVSGNVEPKVNNIKIIPENDIK 1440
1255 ESEKRGORTSTFOJNGKDNKPKIYLKGECLKEISSESVSGNVEPKVNNIKIIPENDIK 1314
1441 SLTYKESAIRFINGDVIEMFPNRRNSSEKSHLSSSDAEGNTRDSLETLPTSKESDST 1500
1315 SLTYKESAIRFINGDVIEMFPNRRNSSEKSHLSSSDAEGNTRDSLETLPTSKESDST 1374
1501 OTTPPSACPPSNNSVNOVEMDEIETSEYKKTYSPTISEESNLSNDFIDNGLPINKNE 1560
1375 OTTPPSACPPSNNSVNOVEMDEIETSEYKKTYSPTISEESNLSNDFIDNGLPINKNE 1434

1561 NVNGESKRVITTEVTMTSTVATESKTIVIKVEGDKQTVVSSYENCAKSTVTTTTT 1620
1435 NVNGESKRVITTEVTMTSTVATESKTIVIKVEGDKQTVVSSYENCAKSTVTTTTT 1494
1621 KLSTPSTGGSVDIISVKEQSKTVVTTVTDLSLTGGTSLVTSMTVSKYSTRDVKLMKF 1660
1495 KLSTPSTGGSVDIISVKEQSKTVVTTVTDLSLTGGTSLVTSMTVSKYSTRDVKLMKF 1554
1681 SRPKKTSRGTALBPSYKRFVMTSTKSTFVLPNDDLKLARKGIRREVYFNYNKPPALDI 1740
1555 SRPKKTSRGTALBPSYKRFVMTSTKSTFVLPNDDLKLARKGIRREVYFNYNKPPALDI 1614
1741 WPPSPRTEGIRTWRYRLQTVKSLAGVSLMRLMLASLRMDMAKAPPGSGSTRTESE 1800
1615 WPPSPRTEGIRTWRYRLQTVKSLAGVSLMRLMLASLRMDMAKAPPGSGSTRTESE 1674
1801 TEITTTTELKRBDVPGYRREYCIKIKICIGVETPKETPTPQKGLNSALRPKRE 1860
1675 TEITTTTELKRBDVPGYRREYCIKIKICIGVETPKETPTPQKGLNSALRPKRE 1734
1861 TPKOTGPVLIETWVAEELLELMEIRAPAEYREKKAQAVEQAKKRLQOKPTVIATSTT 1920
1735 TPKOTGPVLIETWVAEELLELMEIRAPAEYREKKAQAVEQAKKRLQOKPTVIATSTT 1794
1921 SPTSSTSTISPAOKVWVAPISGSVTTGTKMVLTTKVGSPATVTFQONKNFHQTFATWK 1980
1795 SPTSSTSTISPAOKVWVAPISGSVTTGTKMVLTTKVGSPATVTFQONKNFHQTFATWK 1854
1981 QGQNSGVVOYQOQVYLGIIPSTGTSQOFTSEOPRATVYIRNTGSGGTSINSOYIT 2040
1855 QGQNSGVVOYQOQVYLGIIPSTGTSQOFTSEOPRATVYIRNTGSGGTSINSOYIT 1914
2041 GPQIRPGMTVIRTPLOOSTLGLKALIRTPVWYQPAPOVMTQIRGOVPSTAVSAPMTVS 2100
1915 GPQIRPGMTVIRTPLOOSTLGLKALIRTPVWYQPAPOVMTQIRGOVPSTAVSAPMTVS 1974
2101 STPGOKSLTSAFSTSNISQSSASOPPRQOQGVKLTMALQTLQGHGNGGLVIV10GQ 2160
1975 STPGOKSLTSAFSTSNISQSSASOPPRQOQGVKLTMALQTLQGHGNGGLVIV10GQ 2034
2161 OTTGQLOLIPQGVTVLPRPGQLOQAMPNCTVORFLTPLATATATATSTTTVSTTAA 2220
2035 OTTGQLOLIPQGVTVLPRPGQLOQAMPNCTVORFLTPLATATATATSTTTVSTTAA 2094
2221 GTGEORQSKLSPOQVHODKTLPPAOSSSVGPAPAKQFOTAPQAPRPOQTOPQSPAPEV 2280
2095 GTGEORQSKLSPOQVHODKTLPPAOSSSVGPAPAKQFOTAPQAPRPOQTOPQSPAPEV 2154
2281 QTOPEVOTQTTVSSHVDSEAPRTHAQSSKPOVAASOPQOSNVQOSPVVVOQSPSQRIRP 2340
2155 QTOPEVOTQTTVSSHVDSEAPRTHAQSSKPOVAASOPQOSNVQOSPVVVOQSPSQRIRP 2214
2341 STPSQLSPGQOSQVOTTTSTQPIQPHTSLOIPBQGPQOSPOVQOSSSTQTLSSQTLN 2400
2215 STPSQLSPGQOSQVOTTTSTQPIQPHTSLOIPBQGPQOSPOVQOSSSTQTLSSQTLN 2274
2401 SVSSPSRPOLOIQOPQOVAVLVPOLQOOQVQVLSQISQVAAQIOAQOSGVPQOIKLOPI 2460
2275 SVSSPSRPOLOIQOPQOVAVLVPOLQOOQVQVLSQISQVAAQIOAQOSGVPQOIKLOPI 2334
2461 QIQOSSAVQTHQIQONVTVQAASVQEOQLQRYOQLRDQOQKKQOOLEINVTBPSKLLIKV 2520
2335 QIQOSSAVQTHQIQONVTVQAASVQEOQLQRYOQLRDQOQKKQOOLEIKREHTLOASNQ 2394
2521 EITOKOVYMKHNAVIEHLKQKSMTPAREBENQMIYCNQMYKILDKIKKEEQQAARR 2580
2395 EITOKOVYMKHNAVIEHLKQKSMTPAREBENQMIYCNQMYKILDKIKKEEQQAARR 2454
2581 KREESVQOKRSKONATKLSALLFKHKEOLRAETILKKRALLDKDQIOVEBELRDLKIKK 2640
2455 KREESVQOKRSKONATKLSALLFKHKEOLRAETILKKRALLDKDQIOVEBELRDLKIKK 2514

QY 2641 EKDLMLAQAATAVAAACPPPTVLPAPAPAPPPPPPPVQHGLSTPTLPAVASQAKRR 2700
 DB 2515 EKDLMLAQAATAVAAACPPPTVLPAPAPAPPPPPPPVQHGLSTPTLPAVASQAKRR 2574
 QY 2701 EEEKDSSSKKKKKMISTSTKETKDKTKLCTKTPYDESKFYIGDRCQNMWVGRCVGI 2760
 DB 2575 EEEKDSSSKKKKKMISTSTKETKDKTKLCTKTPYDESKFYIGDRCQNMWVGRCVGI 2634
 QY 2761 LQSEAELEIDYVCPQOSTEDAMTVLPLETEKDEGLKRYLRSLOAHKAMPFLPEVDPN 2820
 DB 2635 LQSEAELEIDYVCPQOSTEDAMTVLPLETEKDEGLKRYLRSLOAHKAMPFLPEVDPN 2694
 QY 2821 DADPYGVIKPEPDLATMEERVOARRYEKLTPEVADMTKLFDCRCRYNPSDSPFYQCAEV 2880
 DB 2695 DADPYGVIKPEPDLATMEERVOARRYEKLTPEVADMTKLFDCRCRYNPSDSPFYQCAEV 2754
 QY 2881 LESFVQKLGFKASRSHNNKLOSTAS 2907
 DB 2755 LESFVQKLGFKASRSHNNKLOSTAS 2781

RESULT 3

AAB94078
 ID AAB94078 standard; Protein; 557 AA.

AC AAB94078;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14273.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000BP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PS Claim 8; SEQ ID 14273; 2537BP + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 557 AA;

Query Match 18.0%; Score 2695.5; DB 22; Length 557;
 Best Local Similarity 79.0%; Pred. No. 1,1e-129;
 Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;

QY 1937 MVAPIGSGVTTGKMWLTTRKVGSPATVTFQONKRFHGFATWVKQGSNGVVOQKVL 1996
 DB 1 MVAPIGSGVTTGKMWLTTRKVGSPATVTFQONKRFHGFATWVKQGSNGVVOQKVL 60
 QY 1997 GIIPSSGTSGQTFSTQPTATVTRPNTSGSGGTTSNQVITGPIRPGMTVIRPQLQ 2056
 DB 61 GIIPSSGTSGQTFSTQPTATVTRPNTSGSGGTTSNQVITGPIRPGMTVIRPQLQ 120
 QY 2057 QSTLGRKAIIRTPVWVQPGAPQVMTQIIRGQPYSTAVSAPNTVSSPPGKSLTASTSN 2116
 DB 121 QSTLGRKAIIRTPVWVQPGAPQVMTQIIRGQPYSTAVSAPNTVSSPPGKSLTASTSN 180
 QY 2117 IQSSASQPPRPQGGQVYKLTMAOLTQLTQGHGNGGLTVVLIQGGQGTGQLIPQGVTVL 2176
 DB 181 IQSSASQPPRPQGGQVYKLTMAOLTQLTQGHGNGGLTVVLIQGGQGTGQLIPQGVTVL 240
 QY 2177 PGPGQQLMQAAMNGVYQRLFTPLATTTATTTTSTTAAGGEOQSLSPQMOV 2236
 DB 241 PGPGQQLMQAAMNGVYQRLFTPLATTTATTTTSTTAAGGEOQSLSPQMOV 300
 QY 2237 HODKTLPPAOSSSVGPAAKAPQTAAPSAPQPTQPSAPQPEVQTPQEVQTTVSSHV 2296
 DB 301 HODKTLPPAOSSSVGPAAKAPQTAAPSAPQPTQPSAPQPEVQTPQEVQTTVSSHV 360
 QY 2297 PSEAQPTHAQSSKPYVAAQSQPSNVQGSFVYVQSPQRIIRPSPQSLSPQGSQVQRT 2356
 DB 361 PSEAQPTHAQSSKPYVAAQSQPSNVQGSFVYVQSPQRIIRPSPQSLSPQGSQVQRT 420
 QY 2357 TTSQPIPIQPHSTLSQIPSSQGPQSPQVQVSTQTLSSGQTLNVSVSPRPQLQIQPQ 2416
 DB 421 TTSQPIPIQPHSTLSQIPSSQGPQSPQVQVSTQTLSSGQTLNVSVSPRPQLQIQPQ 446
 QY 2417 PVIYAVPQLQVQVLSIQSQVVAQIQAAQSGVPPQIKQLPQLIQSSAVQTHQIQNV 2476
 DB 447 ----- 446
 QY 2477 VIVQAAVQEOQLQVQQLRDQOQKKQOQTEINWNPSSKLLIKVEIIQKVYKHNNAVIE 2536
 DB 447 -----QVYKHNNAVIE 457
 QY 2537 HIKQKSMTPAEREENQRMIVCNQVKKYIILDKTDKEKQAARKKREESVEQKRSQKAT 2596
 DB 458 HIKQKSMTPAEREENQRMIVCNQVKKYIILDKTDKEKQAARKKREESVEQKRSQKAT 517
 QY 2597 KLSALLFKHKEQRAETLKKRALDLQIEVOEELKRLD 2636
 DB 518 KLSALLFKHKEQRAETLKKRALDLQIEVOEELKRLD 557

RESULT 4
 ABB68735
 ID ABB68735 standard; Protein; 976 AA.

AC ABB68735;

DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 32997.
 XX KM Drosophila: developmental biology; cell signalling; insecticide;
 XX KM pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PMD, Myers EW;
 XX DR WPI: 2001-656860/75.
 XX DR N-PSDB; ABL12838.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 PS Disclosure; SEQ ID NO 32997; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 976 AA;
 Query Match 9.9%; Score 1488.5; DB 22; Length 976;
 Best Local Similarity 31.9%; Pred. No. 1.3e-67;
 Matches 353; Conservative 139; Mismatches 295; Indels 319; Gaps 21;

DB 443 SKILILSLDAEIELRLHSQITERRDEIEROKMLTETLTNEHKHKRSVIT----- 494
 QY 419 ESTAKKGGIDNVKSPETEEKDKNETENDSKDAEKNEFEEDOSLEDSDDKTPDDDDQ 478
 DB 495 -----ETIOEAKNELLKEVLIED--EKDGAKSSESQSI EG 528
 QY 479 GKSEPTPEVDKGNVSANLGDNTTNTATSETPSEGRSPVGCLETPDDSNNAEKVVAS 538
 DB 529 TKRQE----- 533
 QY 539 ELPODVPEERKTCESSTNTATTTSTIOPNLNENSSSELNMSQSESAKADDPENGERES 598
 DB 534 ----- 533
 QY 599 HTPVSIQEIYDPTSEKSTGELSESPGAGKAGSGSTRIITRLRNPDLSLKSOQVAA 658
 DB 534 -----EKMVTR-----OKSNQLT- 547
 QY 659 AANEANKLFKEGKEVLVNSOGELISRLSTKKEVIMKGNINNVKLGQEGKYRYVHNOYST 718
 DB 548 -----NGTLHFKEGMEGFKYVNOYST 570
 QY 719 NSFALNKHOREDDKRRHLAKKCLTPAGEFKNGSVHSGSKVLTITLTALTTOLENNI 778
 DB 571 NPILNKPORNEERDKRRHSHKFSLTASDFKWTGITMTDMMITTLKQTLINFSNI 630
 QY PSSPFHNMWASHRAMVTKAVOMCSPREFALALALICAVKPVVMLPIRREFIGHTRLHR 838
 DB 631 AASFLLNINWVYNNKTIWMAAVNANRPREFAVLLLPASLSKSVFANVNEHQLGHTLQR 630
 QY 839 MTSIEREREKKEVKKKKKKQ--EEETMQATVWYKTPPVKHQVWKQGEYRVYTGQW 895
 DB 691 ITSAEREREKRLKEKREKREDEERNRALFNVIKYTLGLKHQWKKQGEYRVHNGQW 750
 QY 896 SWISKT-----HYRFPKPLBNTNVYKRSBETKNNMDENMDESRRKCSPPKIKI 950
 DB 751 LMLSSSRRCGVARRAPQLHNRYVHYTM--GEENDVNEIT-----LVDPRTQRF 799
 QY 951 EPDEKEKEVGSDAKADQ--NEMDISKITTEKKDQVKEILDSDDKPKKEPMEVDD 1008
 DB 800 MOCCESSNVDCQYCHYLPDQYKNKYIEDYTE----- 831
 QY 1009 MKTESHVNCQESSQDVVNVSEGFHL--RTSYKKTKTKSKLUDILERRIKQFLTEKQRL 1066
 DB 832 -KIKGH-----IDVSKALNAPGRITYYSKARKSRUDLDRLKLAEEVEM-- 877
 QY 1067 EKIKLEGIGIKGKTSTNSKNLSES 1092
 DB 878 -ASKIPSDMKPLLVSSQNNTRANSKQT 902
 RESULT 5
 AAB54364
 ID AAB54364 standard; protein; 328 AA.
 XX AC AAB54364;
 XX DT 09-MAR-2001 (first entry)
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
 XX KM Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KM detection; diagnosis; identification; cytostatic; neuroprotective;
 KM neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KM antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KM linkage analysis; tissue identification; tissue typing; forensic;
 KM neutral; immune system; muscular; reproductive; gastrointestinal;
 KM pulmonary; cardiovascular; renal; proliferative.
 OS Homo sapiens.
 XX OS
 PN WO200055320-A1.

XX 21-SEP-2000.
 PD 08-MAR-2000; 2000MO-US05989.
 XX PF 12-MAR-1999: 990S-0124270.
 XX PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA Rosen CA, Ruben SM;
 XX PI MPI: 2000-579444/54.
 XX DR N-PSDB; AAC99129.
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 PS Claim 11: Page 1272-1273; 1379pp; English.
 XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytosolic,
 CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX Sequence 328 AA;
 SQ
 Query Match 9.3%; Score 1385; DB 21; Length 328;
 Best local Similarity 66.4%; Pred. No. 5.6e-63;
 Matches 286; Conservative 0; Mismatches 17; Indels 128; Gaps 2;

OY 764 ISTRLTITOLENNIPSSFFHPMASHRANWIKAVQCSKPREPALALIECAVPM 823
 DB 246 ISTRLTITOLENNIPSSFFHPMASHRANWIKAVQCSKXREFALALIECAVPM 305
 OY 824 LPTWREFLGHT 834
 DB 306 LPTWREFLGHT 316
 RESULT 6
 AAY04323
 ID AAY04323 standard; Protein; 238 AA.
 XX AAY04323;
 AC 18-JUN-1999 (first entry).
 DT 18-JUN-1999 (first entry).
 XX
 DE Human secreted protein SEQ ID NO:69.
 XX Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS.
 XX Homo sapiens.
 OS
 XX
 PN WO9910363-A1.
 PD 04-MAR-1999.
 XX
 PD 27-AUG-1998; 98MO-US17709.
 PF
 XX 29-AUG-1997; 97US-0056271.
 PR 28-AUG-1997; 97US-0056073.
 PR 29-AUG-1997; 97US-0056247.
 PR 29-AUG-1997; 97US-0056270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
 PI MPI: 1999-190585/16.
 DR
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Disclosure; Page 18; 170pp; English.
 XX
 XX AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
 CC AAY04293 to AAY04321 represent the secreted proteins encoded by the 29
 CC human genes. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions,
 CC e.g. by protein or gene therapy. Also pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new genes. Specific
 CC uses are described for each of the 29 genes, based on which tissues they
 CC are most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
 CC used in the exemplification of the present invention.
 XX
 SQ Sequence 238 AA;
 Query Match 6.6%; Score 982.5; DB 20; Length 238;

Best Local Similarity 78.8%; Pred. No. 1.4e-42;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;

QY 2689 PTLVVASQKRRKEFEK-----DSSSKSKKKMMISTTSKEFEK-----DT 2727
Db 3 PTRP-----KTPYDSKFKYIGCDLCTNMVHGRCVGTTEKAKKMDVYICNDCKRAQESSSE 58
QY 2728 KLYCICKTPYDESKFYIGCDRCQWYHGRCVGILOSSEALIDEVYCPQOSTEDAMTVLT 2787
Db 59 ELYCICRTPYDESQFYIGCDRCQWYHGRCVGILOSSEALIDEVYCPQOSTEDAMTVLT 118
QY 2788 PLTEKDYEGKLRVRSLOAHMAMPLEPVPNDAPDYGVYIKPEMDLATMEERQRRY 2847
Db 119 PLTEKDYEGKLRVRSLOAHMAMPLEPVPNDAPDYGVYIKPEMDLATMEERQRRY 178
QY 2848 EKLEFVADMTKIFDNCRRYVPSDSPFYQCAVELESFPYOKLKGFKASRSHNNKLOSTAS 2907
Db 179 EKLEFVADMTKIFDNCRRYVPSDSPFYQCAVELESFPYOKLKGFKASRSHNNKLOSTAS 238

RESULT 7
AAU19446
ID AAU19446 standard; Protein; 237 AA.
AC AAU19446;
XX
DT 04-DEC-2001 (first entry)
DE
XX Human diagnostic and therapeutic polypeptide (DITHP) #32.
XX
KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder.
XX
OS Homo sapiens.
XX
PN W0200162927-A2.
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001MO-US06059.
XX
PR 24-FEB-2000; 2000US-0184693.
PR 24-FEB-2000; 2000US-0184697.
PR 24-FEB-2000; 2000US-0184698.
PR 24-FEB-2000; 2000US-0184768.
PR 24-FEB-2000; 2000US-0184769.
PR 24-FEB-2000; 2000US-0184770.
PR 24-FEB-2000; 2000US-0184771.
PR 24-FEB-2000; 2000US-0184772.
PR 24-FEB-2000; 2000US-0184773.
PR 24-FEB-2000; 2000US-0184774.
PR 24-FEB-2000; 2000US-0184776.
PR 24-FEB-2000; 2000US-0184777.
PR 24-FEB-2000; 2000US-0184797.
PR 24-FEB-2000; 2000US-0184813.
PR 24-FEB-2000; 2000US-0184837.
PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204225.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.

PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
PA (INCY) INCYTE GENOMICS INC.
XX
PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Ameshey S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiter TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR WPI: 2001-502867/55.
DR N-PSDB; AAS31017.
XX
PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT
PS Claim 27; Page 418-419; 522pp; English.
XX
CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs. By expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
CC therapeutic (DITHP) polypeptides of the invention.
XX
SQ Sequence 237 AA;

Query Match 6.6%; Score 981; DB:22; Length 237;
Best Local Similarity 87.5%; Pred. No. 1.6e-42;
Matches 182; Conservative 4; Mismatches 6; Indels 16; Gaps 1;

QY 2716 ISTTSKETKR-----DTKLYCICKTPYDESKRYTIGCDRCQWYHGRCV 2759
Db 30 VGI TEKEAKKMDVYICNDCKRAQEGSSEELYCICRTPYDESQFYIGCDRCQWYHGRCV 89
QY 2760 ILOSSEALIDEVYCPQOSTEDAMTVLTPTLEDYGLKRVLSLOAHMAMPLEPVP 2819
Db 90 ILOSSEALIDEVYCPQOSTEDAMTVLTPTLEDYGLKRVLSLOAHMAMPLEPVP 149
QY 2820 NDA PDYGVYIKPEMDLATMEERVQRRYRTEKLEFVADMTKIFDNCRRYVPSDSPFYQCAE 2879
Db 150 NDA PDYGVYIKPEMDLATMEERVQRRYRTEKLEFVADMTKIFDNCRRYVPSDSPFYQCAE 209
QY 2880 VLESFVOKLKGFKASRSHNNKLOSTAS 2907
Db 210 VLESFVOKLKGFKASRSHNNKLOSTAS 237

RESULT 8
AAU78845
ID AAU78845 standard; Protein; 149 AA.


```
XX AC AAU78845;
XX DT 18-JUN-2002 (first entry)
XX DE SCAN/KRAB protein SKAT-2 binding protein, 7A1.
XX KM SCAN/KRAB protein associated with a Th2 phenotype; SKAT-2; asthma;
XX KM allergy; hay fever; atopic dermatitis; allergic rhinitis;
XX KM Alzheimer's disease; neurological disorder; apoptosis;
XX KM T-helper response; vaccination; immune response; autoimmune disease;
XX KM infection; cancer; kruppel-like zinc finger protein; 7A1; human.
XX OS Homo sapiens.
XX PN GB2364051-A.
XX PD 16-JAN-2002.
XX PF 06-APR-2000; 2000GB-0008549.
XX PR 06-APR-2000; 2000GB-0008549.
XX PA (GLAXO ) GLAXO GROUP LTD.
XX PI Blanchard AD, Champion B, Page KR;
XX DR WPI; 2002-306794/35.
XX DR N-PSDB: ABK47549.
XX PT New SKAT-2 polypeptide transcription factor, useful for treating or
XX PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
XX PT acid, antibodies and modulators -
XX PS .Disclosure; Page 37-38; 47pp; English.
XX CC The invention describes an isolated SCAN/KRAB protein associated with a
XX CC Th2 phenotype (SKAT-2) polypeptide (I). (I), especially where expressed
XX CC in recombinant cells, is used to identify agents that modulate SKAT-2
XX CC activity, and may be useful therapeutically. The agents, also (I) and
XX CC the polynucleotides (II) that encode it, are useful for treatment and
XX CC prevention of diseases responsive to SKAT-2 modulation, specifically
XX CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
XX CC Alzheimer's disease (and other neurological disorders, particularly where
XX CC related to apoptosis); conditions involving T-helper responses, and
XX CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
XX CC or humoral responses, to reduce development of a Th1 phenotype, and to
XX CC manipulate the immune response in autoimmune diseases, infections and
XX CC cancer. (II) is also used for recombinant production of (I); as source
XX CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
XX CC detection of mutations and for monitoring SKAT-2 expression in
XX CC association with disease. This is the amino acid sequence of the peptide
XX CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
XX CC SKAT-2.
XX SQ
XX Sequence 149 AA;
SQ
Query Match 5.3%; Score 789; DB 23; Length 149;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 9
AAU31944
ID AAU31944 standard; Protein; 246 AA.
XX AC AAU31944;
XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #2435.
XX KM Human; vaccination; gene therapy; nutritional supplement;
XX KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US08656.
XX PR 16-APR-2000; 2000US-0552929.
XX PR 26-JAN-2001; 2001US-0770160.
XX PA (HYSE-) HYSBO INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20; Page 529-530; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising
XX CC the nucleic acids encoding the polypeptides and cells genetically
XX CC engineered to express them are also useful for producing the proteins.
XX CC The proteins are useful in genetic vaccination, testing and
XX CC therapy, and can be used as nutritional supplements. They may be used to
XX CC increase stem cell proliferation; to regulate haematopoiesis; and in
XX CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX CC immune suppression and/or stimulation; as anti-inflammatory agents; and
XX CC in treatment of leukemias. AAU29510-AAU33304 represent the amino acid
XX CC sequences of novel human secreted proteins of the invention.
XX SQ
XX Sequence 246 AA;
SQ
Query Match 5.2%; Score 781; DB 22; Length 246;
Best Local Similarity 67.8%; Pred. No. 2.9e-32;
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;
```


XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12500.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB101840-AB1616175), expressed DNA
XX sequences (AB101840-AB1616175) and the encoded proteins
XX (AB857737-AB872072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2768 AA:
Query Match 4.28; Score 621.5; DB 22; Length 2768;
Best Local Similarity 18.9%; Pred. No. 1.2e-22;
Matches 584; Conservative 448; Mismatches 1197; Indels 865; Gaps 133;
OY 78 PLEEKDIP---PLEPKSSEDLAVPNEHIMNVAIYEVLNREGTVLRISFREDCA 134
DB 84 PITTCEPVADVPHSPETELSIPEK---GCSIEKFPYPSGAQVPSNPKFCCLCYC 140
OY 135 LVSOEOSTLMAENHVLKAVLREEDTSNTTFGPADLKDVSNTLYFIDGTMPEVLRV 194
DB 141 INNOTKCVMO-----ECGLH VDG-----CLPIY 163
OY 195 CESDKENHVLPOEADYRPGVENKIKVLOFLVDOFLTN----- 236
DB 164 NKGS-----CCPYRSCDH-----ENELD-----FMDSTTTTTPVPTGFLASTMP 209
OY 237 -----IARELMSEGV-IQYDHCRCVCHL-GDLLC-----CE-----TCSAYYHL 275
DB 210 PTTTDCIHGEITPADGSLGKNACEHCYCMRGDVCAYOCECEPMAANGKSCRAPAA 269
OY 276 ECVKPLPEEVEDMOCQEVCAHKVPGVTDV-----AETOKKPYIRHEPIGYDSRR 329
DB 270 EGGCCSNVYCEDSSSTTELVETTPESATSVPAKGHAIEPKEDVLOEHIDDDKNKE 329
OY 330 KYFPLNRLIIEEDTENENKKTWYSTKVOLAEIDCLDKYW---EAEICKILEMRE 386
DB 330 TATIPSAELGSEIEEEEEKD---KATVAPVQVTD--EKDFSEPESSTAGIPDSR- 383
OY 387 EHRHMDITEDLNKAGSKSFLAANEELISIRAKK---GDIDNVSPETEEDKN 442
DB 384 -----IDLPSSTEEKSESTEAEDIVKIVTTPPEPGSGEEDVPKPSQIPEKEIT 434

OY 443 ETE-----NDSKDAEKNEE-----FEDQ-----SLEKSDDKTPDDDEPG 479
DB 435 EDELIKVTSAPAKASPEEEVYATTSAPTEEDVKPTAGTISEEEBGRKPTP---AEEG 491
OY 480 KSEF-----PTEVGDKNSVSNLGDNTTATSEETSPSEGRS-----P 518
DB 492 SGEEEKDVKYTAPEETEDEKPTSAVY---ASDEKDEQPRPSGSGDELDLKPPTAPT 548
OY 519 VGCESETPDSSNMAEKKVASELPQDQPE-BENKCCESSTNTSATTSIOPNLNENSSSEL 577
DB 549 AGATSASESEEDQDEBK-STPAPTSVDIDIEPAKPTESSEASG-----EGEDVAKET 599
OY 578 NSSQESAKAADDPENGERESHPTVSI-QEEIV-GDPTSEKSTGELSPPAGKAGAGST 635
DB 600 TPAGEASLAGEEETIVKGTTPAGEPSSGEDEEIVKGTTPAESESESEDE----- 648
OY 636 RIITRLNPNDSKLSQLSQVAAAHENKLFKRGKE--VLVYVSGQISLSTKKEVIM 693
DB 649 --LTKVTTPAGEPSPVAGEEETIAKETTTPAGEPSIAGEEIVVTTPAGE--SSIAGEEIVK 705
OY 694 KGINNFKLGOEGKYRVYHNOYSTNSPALNKHOREDDHRRHLAHFCLTPAGEEFKMN 753
DB 706 -----VTTPAGESSSGEEELIK-----VTTPAGESSSE 734
OY 754 GSVHGSKYLT-----ISTRLTITOLENNITPSPFHPNMAHRANWIK--VOMCSKPRE 806
DB 735 GDEEIVKESTPAGEPSISGEEDVIKATTSAPKSDIEGVKEPETATEVPAEEVEDPAK- 792
OY 807 FALALILECAVKNVVMPIWREFLGHRLHRMTSIEBEKEKKKKKEKOEETMOQA 866
DB 793 ---TTPAIAEAEPEEPIAGTPIPTDGIS-----GEEIVKGTTPOTLOEPEISES 839
OY 867 TWVYTPPVKKQVWKQGEERYVTGYGWSVISTHYRVFKIPGNTNVN-YKRSLEGT 925
DB 840 TEV-----PVAED-----DLSSSTSASASTEGV 865
OY 926 KNNMDENMDESCKRCKSPKIKIIEPDEKD-----EYKSGDAKAGADON 971
DB 866 QDAASETTSSAPARAGKDEAATVPTPAOKDDEVEDQDADLPEVDVQSTAKTTTTE 925
OY 972 EMDISKITEKKDQVKELSDSDPKCKEPEVEVDDMK----- 1010
DB 926 QPKESSTEAEDAEIEVTTSSPADKQVEPAEPADKHKKDEEDVQATDLPKSDIGBPV 985
OY 1011 --TESHVNCOESSQVDVNVNSEGPHL-RTSYKKRTKSKDLGLLEIRIKOFTL-----E 1061
DB 986 VDTAATGQPTSDETATDKPVSYLEPVSOEVSSTAKVNVNRNDETEKFTLPPSGEDQ 1045
OY 1062 EKORLEKIKLEGIGK-----IGKSTNSSKNLSESPVITKAKEGCOSDSMRQDSFNA 1115
DB 1046 SSEPLPAMDLPAGIPRGDDCLVEGKTYANNITVATPACDVSK--CISLIVACQOMBECK 1103
OY 1116 NNDOPE-----DLQCSQS---DSSVLRMSDPSHTTNKIKYPRDVLVDVYSISPEKIC 1166
DB 1104 LPEMLKCTVAAADLLDCCPTTICDESTESAKEDEESTAK--PONKIDEDVEISTE-EI 1160
OY 1167 PKNSIENDIEEK-----VSDLASRGOEPTKSKTKGNDFFIDSKLASADDI 1213
DB 1161 PKDVIPTGITEQPLSHVKPDEELIQVTSYPAQDESTAKVADKP--IDBS----- 1210
OY 1214 GTLICRKNKPLIOESDPTIVSSSKSALHSVSKSTNRDAPTLSRANDFEKLGCDSESN 1273
DB 1211 ---AEDKKPIGESEED-----SKPIDESEEDKKPVE----- 1238
OY 1274 STLENSDVIYSIQDSSEDMIV-----QNSNSESIOQFTRQD-----VEVLPELKE 1322
DB 1239 ---ESAEDKKRVEDSEKEKPLPTVIPASETEKSEKBEDEKKTADPAAPTEQPEATTPA 1295
OY 1323 LVSGESTGNCEDRLPVKGTANGK--KPSQOKKLEBRPVKKGSDOI---KLKVTYTKN 1376
DB 1296 QIADTAEVDDKLATTISAPVSGEDELKPADKKRTE-----TAQIPDAEITPASTDEPE 1349

QY 1377 NENR-----ESEKKQRTST--FOINGKDNKKPIYIKGECLEKIESSRVVSGNVEBKVNNI 1430
 Db 1350 STEELPTVDLKKPEEDSTKGTAEPSDKVPEVPTASTENEIEESDFTTVAAPKISAS 1409
 QY 1431 NKIIP--ENDIKSLTVE-----SAIRPFIINGDI-----MEDNERNSSSTKSL 1474
 Db 1410 DETEPTAEEDLVPAFTFEPIESEFEVSTKPPAVOGPPLPTLAPQDEKKPVDAETSTADI 1469
 QY 1475 LSSSDAEGNRYDLSLTLPSTKESDSTQTTTP--SASCPSNSVNOYEDMEIEFSEVKKYTS 1533
 Db 1470 STEPSAEVEKEASGETSSDNEIDAGASTTVPVPSADDDKPISTKTYE--ADCKFTTV 1526
 QY 1534 SPIT--SEESN--LSNPFIDENGLPT-----NKNENNGESK-----RKT 1570
 Db 1527 APADDEESNLPKLPDPIFEESA--PVAVTTAAPSKD--GEOKPVEVEEKPIEDGCKP 1582
 QY 1571 VITEVTTSTVATSESKIVIKVEKGDQTVVSTENCAKSTVTTTTVTVKLSTSTSGS 1630
 Db 1583 IEDETSTPS-----SENEIPESTRATTIAPSKEE--PSEPSTGAPTKDEPAEPTDAP 1635
 QY 1631 VDIISVKEQSTVTVTTVTDLSLTGTGLVTSMTVSKESTRDVKLKKFSRPKTRSGT 1690
 Db 1636 ESDES--KETPESEVPTTVA--GEKIPTSITPDEPTATSAVPAKPDDEVKEST 1690
 QY 1691 ALPSVRKVTSTKKSIFVLNDDOLKILARKGIREVYFVYNNAKPALDIMPYSPRPTF 1750
 Db 1691 ELPDPAASSEEDENS--STDQIPS-----EVP--EKRP----- 1720
 QY 1751 GITWRKRLQTVKSLAGVSLMLRLMASLRMDMAKVPDGG--GSTRTSETETITTE 1807
 Db 1721 -----ETPAQTPPEBDIYGATAPRTSDVPPVQ 1749
 QY 1808 IIKRDGVPYIREYCIRKILICIGVETPKETPTPQK--GLRSSALRPKRPETPKQTG 1866
 Db 1750 RLPE-----EVLAEIPOPSTETGIRKODETTAAPSIDRK-- 1783
 QY 1867 PVILETVAEELEIMEIRAFAEVREK--EKAQAVEQ-----QAKKRIEQQKPTVIAST 1919
 Db 1784 -----EPVTELEDEBATVAPRISEKDEKPTEEKPEVQKPGEESEEEKKPIEDQVST 1839
 QY 1920 TSPVSTSTSTISPAQKVMVAPISGSVTTGTKMVLTTKVSFATVTFQONKNFHOTFAMV 1979
 Db 1840 EGPVSTSEAGSTES-----SEEVKPSTEGEVAEKPED----- 1873
 QY 1980 KQGGSNS--GVYVOQKVLGIIIPSTG---TSQOTFTSGQPTATVTR--PNTSGSGGTT 2034
 Db 1874 KQPSSTAQAPVETIPEISTELPQADGDKPTSEAPVDSDEDTAPSDEKIPSVSG----- 1927
 QY 2035 NSQVITGPQIRPGMTVIRTPLQOSTLGKATIRTPVMVOPGAPQOV--MQOIIROGPVSTAV 2033
 Db 1928 --EEVEGPEV-----TTASFOAAEDELKTPAESEPSSTDKVPELEYOKPEDETKAD 1977
 QY 2094 SAPNTVSTPGQKSLTSATST--SNIOSASQPPPOGQVKLMAOLTLOLTOGHG 2148
 Db 1978 ETPESTVQVSDVATSTSAVPAVGDIKDEQATASP--EEBEETKPIIAAELIPQ----- 2031
 QY 2149 NQGITVVIQOGQGTGLOLIPQGYVLP--GEGQOLMO--AAMPNGTVQRFLEPTLATTA 2205
 Db 2032 -----PSEKEPEVDE--QEVESGTKATPAESGQPIDEIAPATSGIDE-----ASTA 2076
 QY 2206 TTAATTTTTVSTTAAGTGEQROSKLSPOQVHODK-----TLPPAQSSSVGAPKAPQOT 2259
 Db 2077 APKKEESTVASA-----SP--AVHDEIKDVTTTQPVADKEKVAAPQDDTKT 2123
 QY 2260 A-----QPSAR--POPOTOPQSPAPEVQTOPEVOT--QTVVSHNPS-----EAQPTH- 2304
 Db 2124 SIDVSTDSPTAODKODKTEAPVAPPTVSSPTADSAASTPVEKPSVEIDTKRMD 2183
 QY 2305 --AQSXSPQVA--AQSOPQSNVQGSVPYRV--QSPSOTRIAPS-----TPSQLSFGQOQSO 2353
 Db 2184 IMSOTIAPHADGAASTSTDEDDQAPVTVSPDAEKTVPSPAPQDSKTPSSAPQDAD 2243
 QY 2354 VQTTTSG-----PIPIQPHTSIQIPSGQPGQPOVQSTQTL-----SSQOTLNQVS 2401

Db 2244 IPATATPLDDNKKIPATVAPQTDGVPATAAPALDEDEKIQTTAAPALDEEKIPSTAAALDDEK 2303
 QY 2402 VSSPSRQLOIQQPPQRYIANVQLOQYQVLSQIOSQVVAQIOAQOQSGVPPQIKLOLPQ 2461
 Db 2304 IPAPVSPVFLVEPSESEKPAVSEYDGE-----ESTEPVHVHENTSDPTSDAKLKPT 2358
 QY 2462 IQOSSAVQTHQIQQNVVTVQAA-----SVQEQLOQRVQOLRDQOQKKKQOQIEINVTPEKL 2516
 Db 2359 APATSESPATIEAIVPETTAPELEKEVEPKATQOPELEKETPEKATBOPELEKETPEKA 2418
 QY 2517 LIKVEI-----IQOVYMKHNAVIEHLKOKMSMPAER-----EENQRMIV 2557
 Db 2419 TEQPELEKETPEKATEQPELEKEVTDKATQEPESVDEKTTPEPVKPSLDSTDEDE----- 2474
 QY 2558 CNQWAKYITLDIKDEKKAARKKRESEVEOKRSQONATKLSAL----- 2601
 Db 2475 -----ESVESEESADKDKKKEDEDTDKKHEPEVAVVSEIQQSEAVPTTG 2525
 QY 2602 --LFKKKEQLAEILKRRALLDKDQIEVQELKRDLLKIKKEKDMOLAQATAVAAAPCP 2659
 Db 2526 HPLPPH--LASSITTPAVDR-----VGE-----DEENTVYKLSSTTTSTTESP 2570
 QY 2660 VT-----PVLPAAPAPPSPPPPGVQHTGLSTPLPVAQKRRKREKDSKSKKKM 2715
 Db 2571 VTSAPSTTVASQOQOQPIPTPPYG--HA-----PEYEDEVEDEEYFGGTCRYAGKL 2620
 QY 2716 ISTSKETKDKTKLYCICKTKPYDESKFYIGDRCQNWYHGVCVILDS-----EHEL 2767
 Db 2621 YVSAQOILPRDDPCDRCFC-----FRSDITLQOQSCPPPIAGCHEEP 2661
 QY 2768 IDEVYCPDQ-----STEDAMTVLTP 2788
 Db 2662 ISGFCPRPECPVSMAYLNTTSTTTTSTLTP 2695
 RESULT 12
 ABB65772
 ID ABB65772 standard; Protein: 5533 AA.
 XX
 AC ABB65772;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide seq ID NO 24108.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS
 OS Drosophila melanogaster.
 XX
 PN MO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-0509231.
 PR 23-MAR-2000; 2000OVS-191637P.
 PR 11-JUL-2000; 2000OVS-0614150.
 XX
 PA (PEKE) PE CORP NY.
 PI
 PI Venter JC, Adams M, Li PMD, Myers EM;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL09875.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PS interactions -
 PS
 PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5533 AA:

Query Match 4.1%; Score 613.5; DB 22; Length 5533;
Best local similarity 19.5%; Pred. No. 8.2e-22;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
QY 305 DCVAETIQKPKYIRHEPIGYDRSRKYWFLENR-----RIIIEEDT 344
DB 1912 DTADKAEKKNQ---RHEKEKKEKREKDLKQVREKREKKAQOEEREKEDKAKKEEK 1968
QY 345 ENENKKIMWYTSKYQALAEIDCLDMDYWAELCKIIEKREIHRMDTE-DLTKAR 403
DB 1969 ERERREKKAQEDREK-----KERERELEK-KEGRDKEQKEKEIREKDLREKQ 2015
QY 404 GSNKSEFLAANEELISIRAKKGDINVKSPETEKKNTENDSKDAEKREEFEDQSL 463
DB 2016 -----RERONREKELRKDKLREKEMRKREKREKELHEKQOREKREHR 2057
QY 464 EKOSDKTPDDDE--QGSSEPT-----VGDGNSVSNALGDNNTATSEETSPS 513
DB 2058 EKROSRAMDEQEGGRRELISYOKSKMDIAGEASLTALDCHKNENAMDTIAGT 2117
QY 514 EGSPVGCLETPDSSNMAEKKVASELPDVPREPKN---TCSSNNTSA-----TTSI 564
DB 2118 PGASP-----STP--SDNTPKERSKLSRNSPVRLHRRILSDEPNSAGCGSGGSSH 2170
QY 565 QPNLENSNSSSELNSSQESAKADDPENGERS--HTPVSIQETVGDFTSEKSTELS 622
DB 2171 QIHEDYVVRIRRENSQINISVHSSNQLNDRDRSKHKSSPFED-----KNSSSHIS 2223
QY 623 ESGAGKAGSGSTRITRLRNPDSKLSQLKSQVAAAANFANKLFKEGKVLVYNSOGET 662
DB 2224 RPHGCGSSASSSKHNHR-----RDKHHQKGSASITNSI-----EVVVDPISQTK 2271
QY 683 SRLSTKEVYMKGNINNYFKLQEGKYRVYHNYSTNPFALK-----HQHR----- 729
DB 2272 HNLNTSEELQSHQ-----PAREKREHFSSHANSSSRHRSKRDHHRREKRRHSA 2324
QY 730 -----EDHDKRRHLAKFCILT---PAGEFKWNGSVHSGKVLITSLRLITOLENNIP 779
DB 2325 ESTNTDEEHTPOOHNRHRTISAAGSGSAGELSSAATNTSGKLHHQHRHSRERKSRGS 2384
QY 780 SSGFFHNMASHRANWIKAVOMCSKPREPALALALECAVAVVMLPIWFEFLGHTRLHRM 839
DB 2385 DEGHSSSSKSLRAKLMMLSSADSDDTDA-----SKKHSI 2419
QY 840 TSTERE-----EKEKYKKKEKKEEETMQOATVAVKTPPVKHOVKKQKEEYRVGYCG 894
DB 2420 FDLPDPCPNVSMYDKVAKRSCKMORAEKKIKAKFS-----OLQOSRAKKKRSTYDG 2474
QY 895 WS---WISKTHYVRFYPLKPGNTNVYRKSLSEGTKNMMDENMDESKRKSRSRPKIKIE 951
DB 2475 DSDTEEDRHO-----RNSGSSSFHGR--YFGLSSDDDDDEETIORRS----- 2517
QY 952 PDSEKDEVGSDAKAGADONE-----MDISKTEKKDQVKELLDSDSKPCKEPEPVED 1006
DB 2518 -----SDSDAEHGGOOGASTLADANRVROMO-QNLRRLCDGDS---SED--EIR 2563
QY 1007 DDMKTESHVNCQSSQVDVYVNSEG-----FHLRTSY-----KKTTSKSLDGLL 1051
DB 2564 RNVMKSHFEGKRNSTRIASDESOSOPAPDLITKOEHPIDAPQETKREOLSDSEOKFK 2623

QY 1052 ERRIKOFTLEKORLEKIKLEGIGIKGTSTNSSKNLSESPVITYAKEGCQSDSMROEQ 1111
DB 2624 SRHDSNSIBER-----KLKTEREI-----KTLELGFYNSSEYTGKLEKYESPTERRKHK 2675
QY 1112 SPANANDQPEDLLQGCQSQSSS-----VLRMSDP-----HTNKLYPKRVLDVSI-- 1159
DB 2676 SKRR-----LKSSTADTSAAGPPLVMTPLPLPSIFDVHSSSECKTKDFNDPDLKTEC 2727
QY 1160 -----RSPPTCKPKONSIENDIEEKVSDLASGRQEPPTKSGNDPFIDSK 1206
DB 2728 SSIPLEISAGERRKK 2773
QY 1207 LASADDIGTLICRKNKPLIOESDPTIVSSSKSLHSSPKSTNDROATPLSR----- 1258
DB 2774 KLSKEEHRLLKSKSKSMNSCMTKLINSGA-H---PSTPSLAPTPPSASTAQTSK 2829
QY 1259 ---AMDPEKLGCDSSNSTLSSDPTVSIQDSSEDMIVONSNEISQDFTRQDVE 1314
DB 2830 RGEDKMEFIGIISDEEESQFPPEQAEFNKDIIPSS-----VSTTGPIVSALQTYKQ--- 2881
QY 1315 VLEPILKCELVSGSTGCEDRLPYKGTENGK-----KPSQOKLEBPVKKCSQDIKLN 1370
DB 2882 --EP-----STPVS-----KNEAHITLTHEPEQOQDLER--SRLSGSSSS 2921
QY 1371 TTDKKNNESESEK---KGQRTSTFQINGKDNKPKIYLKECLKEISESRVSGVNEPK 1426
DB 2922 HADREHRRRRREKREKREKQOREQNOIHKSSK-----YETK 2959
QY 1427 VNNINKIIPENDIKSLTVKESAIRPPIINGVIMEDFNERNSSSEKSHLSSDAEGNRD 1486
DB 2960 VDDDNSVDMDEAGRALLEAQ-----LMSDFPKPISE----- 2990
QY 1487 SLETLPTSK---ESDSQTTPASCPESNSVNOV-----EDMEIETSEKVKYTSPT 1537
DB 2991 --EATPSTAAITYRSDMTDVRFSDN--EDNNSVDMTKGVASQEOEOKSKDKKKKKRSK 3047
QY 1538 SEESNLSNDFIDENGLPIKNKENVNGESKRYTTEVTMTSTVATSESTVIKVEGDK 1597
DB 3048 EEKQEKLL-----QQORRESLPVASTSSAPRPGIKLVAVQAASK 3088
QY 1598 QT-----VYSTENCAKSTV-----TTTTTVTKLSPRSGSVDIISVKEQS 1640
DB 3089 HADLQIDAKHISPPYCKPSPSLPCLIGDDDDALHTPKAKPTTPSSRGNDGLTPSRKRP 3148
QY 1641 KTV-----VTTVTYDLSL-----TTGGVLVTS----- 1662
DB 3149 RLISPIKPTLIANSSTLSTQSAETPVSSGTVISSSLATTPISSTAGYSAAPGLDNP 3208
QY 1663 -----MTVSEKYST-----RDKVYLMKFSRPRK 1685
DB 3209 TSASAQCKKKESEFIPGHDGLDRIRISAVQISIAENSTSLDNIINDEKIPVAPSPRA 3268
QY 1686 TRSGTALPSTRKTVYVSTK-----SI 1707
DB 3269 TKPLDKLEESKSRVTISQETTESAVSALGSEFQSTSTDYSLDGMEMSVNELEPTL 3328
QY 1708 FVLPNDDLKTLARK-----GIREVPYFNNAKALDIWPPSP-----RPTGKIT 1753
DB 3329 VIAEPDEAALAKAIETTAGEPASILEP-----EMEREPAPDPDEALESEPVVEVL 3384
QY 1754 WRYRL-QTVKSLAGVSLMLLWASLRMDMAAKVPPGGSTFTESETEITTELIKRR 1812
DB 3385 DPRELKNAVQSLKHEDWM-----DIKADTPQSEHDQIDT--DTEENDE---AD 3429
QY 1813 DVGPGYIGRFEYCIRKIIICPGVPEPKETPTPOKGLRSSALRPKRPEPKQGVPIIET 1872
DB 3430 SSGP-----SLKIDETVOSSSPPEKSISSNS-----PTPRETANIDI-- 3466
QY 1873 WVAEELELWEIRAFARVRYEKEKAQAVEQAQAKKRLBOCK-----TVIATSTTSPTST 1926
DB 3467 -----PNVESQPKLSMESTPQPSVITKLPLFLDTPKTVAGLPPSPVKIE 3510

```

QY 1927 TSTISPAQKVVAVPISGVS-----TTGKMLVLTTKVSGSPATVFOONKFNQTEATWVKOG 1982
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3511 PRTISKLDQPLVGVVGLVLPARHSTGSGISANSVINIDLNSVSSCNTASASTASATAASA 3570
QY 1983 QSNNGVVOVQOKVLGITPSSITSGQFTFSQP-RTATVITRPVSSGGTSSN---SQV 2038
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3571 SISFSPASQNMAM---POASTPKQGPITPQOALRTQSLIMQPTISIPEDTPHFVAPOM 3627
QY 2039 ITGFOI-----RPG--MTVIRPLQOSTL---GKAIIRT-----PVMVPGAPQ 2078
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3628 VLSQSHHPQDPGTVMGILAPRSPHSLHSPGKGVAAQSLVGLSLPVGKRPVWSQP--SPQ 3666
QY 2079 VMDQILNGQVSTAVSAPNTVYSTPGQSLTSATSTSNIOSA---SQPRPQOGQVKLT 2135
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3687 ---QVOOTQOQNALITRQSSNISPLASPTTRVLSSSNSPTSKVNSYQPRNQ----- 3737
QY 2136 MAQTLQTLQHGNGQGLTVVYIGQGGOTTGOLILPGQTVLPRGGQOLMOAAMPNGTVQR 2195
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3738 -----VPOQRPKSVAEVQTPQOLMTI---PLQKMTPTQVPHNPT-- 3774
QY 2196 FLTPPLATATATATTTTSTTAASTAGTGEOQSKLSPOMQNOCKTLPRQSSSVGRAPA 2255
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3775 -IISKVTVTQVQOATQSOQVASSPRLGS-----LPRHKNVH---LMAON-----Q 3816
QY 2256 QPQT-AQPSARROPOTQPSAPQREVOVOTQOTTVSSHVPEAQRTNAQSSKPOVAA 2314
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3817 QPVYAKMTAHNOHQNHQOQFNMQIOQOQNMQ-----QOQLHQSQS--QITS 3862
QY 2315 QSQQSVNQSQSVRVQVSPQSTKTRSTPQSLSPQ--QSQVQTTTSPQRTQPTSLQIP 2373
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3863 APQOMQIQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3918
QY 2374 SQGQSQSQSQSVQVS---STQTLSSGQTLNQVSSSPSRPQDQIQOPQ--QVLAIVPQLQOQ- 2428
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3919 -QHGVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3973
QY 2429 -----VOVLSQISOQVVAQIOAQOQSVRQOQIKQDLPQIOQSSAVQTNQIQNV 2476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3974 PQGVHLSGSTSIFASQOQNSQLPAR-----GVPOQ---QHPQQLSHSPCKPNTLVSV 4024
QY 2477 -VTVQASVQEQDQOQVQOQDQOQKQKQO-----QEIINV--NTP 2513
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4025 NQGVQRPAILTRVGSHPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 4084
QY 2514 SKLIIKEIIT-OKQV---VMKHNVAIEHLKQKSMTPAREENQRMIVQNMVKYITDKI 2569
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4085 -KIIVQOHIYAQNOVPRPQOQNAIHPQOQKQSTPRGH----- 4123
QY 2570 DKEKQAKKRRKRESEVQKRSKONATKLSAL-----LFKHEQQLPAETLKRALLDK 2622
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4124 -VEFTPMSAQKTESVSVIRPTPTGLAVISANTVGSILTEENLIKISQKODELIEQ 4182
QY 2623 DLOEQV-----BELKRDJKIKK-----EKDMLQALQATAVAAQRPVTPVLPAPRAP 2670
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4183 DSK-EVUSDVWSAKEVINDVYIKLDTPLASKDKAKRAVEQAL-----AP 4226
QY 2671 PPSPRPVQVQHTGLSTPLPVAQKRRKEEKSSSKSKKKM 2715
Db 4227 APIPNQPGNQ--SMAQETALPPTSMVSNNSNDHDEDETETRL 4269

```

```

XX OS Drosophila melanogaster.
XX PN NO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PK 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW,
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL15263.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS Disclosure; SEQ ID NO 40272, 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 5560 AA:
Query Match 4.1%; Score 613.5; DB 22; Length 5560;
Best Local Similarity 19.5%; Pred. No. 8.2e-22;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
QY 305 DCVAEIQKPKYIENHPIGYDRSRKKYWFLLNR-----RLIEEDT 344
Db 1912 DTADKAEKNQ---RHEKKEKQREKDLKQYREBKDRKAQOERREKDRKAEK 1968
QY 345 ENENEKTIWYISTVQVLAELIDCDKYWEALCKIIEBKAREELHNMDETDLTAKAR 403
Db 1969 EREREKRAQEDREK-----KERERELE--KEQDKQKEREIREKDLREKEQ 2015
QY 404 GSNKSFLLAANEELIESIRAKKGDIDVKSPEFEKQNETENDSKDAEKREFEQSL 463
Db 2016 -----RENDNREKELRDKDLREKEREKEREKELHNRKQDREKRN 2057
QY 464 EKUSDXTPPDDPE--QKSEEPTE-----VGDKGNSVANLGNNTVANSSETPS 513
Db 2058 EKEQSRAMVPEQEGRGGRMEELSSYQSKMDIDGEEASSLTAIDCQNKENAMDTIAQGT 2117
QY 514 EGRSPVGCLESTPDPSSMAEKKVASELPQVYREPKN---TCESSTNSA-----TTSTI 564
Db 2118 PGASP-----STP--SDNTPKERSRKLRSNPVHLHRRLLSSQESNHSAGGSGGSSH 2170
QY 565 QPNLENSNSSSELNQSQSESAKAADPENGERRS--HTPVSIQDEIYGDFTSEKSTGELS 622
Db 2171 QIHIEDYVKRIRAMNSQNIYSVHSSNQRLNDRDSKEKKSSEFKD-----KNSSSHIS 2223
QY 623 ESPGAGGAGSGSTRITRLRNPDKLSQLKSOQVAAAHANKLFKEGKVLVNSQGEI 682
Db 2224 RPHCGGSSGSSASSKHNHR-----RDKHNQKGSASSIETNSI-----EVVVDPIQTK 2271
QY 683 SRLSTKKEVIMKGININNYFKLQEGKRVYVHNQYSTSPALNK-----HQRH----- 729

```

```

RESULT 13
ABBT1160
ID ABBT1160 standard; Protein; 5560 AA.
XX ABBT1160;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 40272.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

```

Db 2272 HNLNTSEELQSHO-----PKREKEREHSSHSANSSSRHKS KDHHHREKRRHSA 2324
 QY 730 -----EDHDKRRHLAKHFCILT---PAGEFKWNGSVHSGKVLITLITLRLITPOLENNIP 779
 Db 2325 ESTNTDEHPTPOOHNPRIIRISIAAGSGSAGELSSAATNTSSGKLHGHHRHSVEKSSRG 2364
 QY 780 SSFFHPHMAHRAWIKAVOMCSKPREPALALALIECAVPRVWMLPWREFLGHTRLHRM 839
 Db 2385 DEGHSSSKSLRAKLMMLSSADSDTDDA-----SKKHSI 2419
 QY 840 TSIERE-----EKEVYKKKKKOEETMQOATWVKYTFPVKHOVMKOKGEBYAVTGYG 894
 Db 2420 FDIPLDDCPNVSMTDKYARSCKNMORAEKIKAFS-----OLKOSRAKKKSTSYDG 2474
 QY 895 WS---WISKTHVYREVPKLPNTVNYRKSLDEGTKNMMDENMDESDRKCSRSPPKIKIE 951
 Db 2475 DSDPEFEDRQH-----RNGSSSFHGR--YPLGSSDDDDDETHRRIS----- 2517
 QY 952 PDSEKDEVKSDAKADONE-----MDISKITEKDQYKELDSDSKPCKEEPMVD 1006
 Db 2518 -----SDSDAHEGODNOGASTLADANRVROMQ--QNLRLCDGDS--SED--EIR 2563
 QY 1007 DDKTSHVANCSSQYDVYVNVSEG-----FHLRYSY-----KKKTKSKLDGL 1051
 Db 2564 RYWKHSHFKRKSNSSTRIRASDESOSQAPADLTIKOEHPIAPAQETIKREOLSDSECKFK 2623
 QY 1052 ERRIKOFTLEKORLEKIKLEGIGIKGTSTNSKMLSESPLYTKAKEGOSDSMRQO 1111
 Db 2624 SRHDSNSIER-----KLKTEREI---KTELGDYNSSETTYGKLEKESPEIRKKNHK 2675
 QY 1112 SPANANDQPEDLLOGCOSSS-----VLRMSDP-----HTTNKLYPKDRLDVS1-- 1159
 Db 2676 SKRR-----LKSSSTAOTSAQPLVMTPLTPSIFDVHSSECKTKEDNFDLKTCC 2727
 QY 1160 -----RSPETCKPCKNSIENDIEEKYSDLASGOEPTKCTKGNDFIDSK 1206
 Db 2728 SSTPLETSAEBRKKHKKRKKREKRLKNNTEATYPN-----SPTTMDT-----SSE 2773
 QY 1207 LASADIGLTLCKNKKPLIOESDPIVSSSKSALHSSVPKSTNDRTATPLSR----- 1258
 Db 2774 KLSKEERHLRKKSKKSMONSCWTKLYNSSGA-H---PSTPSLPATPTPSAPSTAOTSK 2829
 QY 1259 -----AMFEKLGCDSESNSTLNSDPIVSIODSEEDMAYONSNESTSEOFRTREDVE 1314
 Db 2830 RGEDKMEFIIIGLISDEESOFPEQAEITNKDIIPISS-----VSTGPIVASALQTYKO--- 2881
 QY 1315 VLEPLKELVSGESTGCEORLPYKGTENGK---KPSOOKLEERPVNKKSOQILKN 1370
 Db 2882 --EP-----STPVS-----KNEFAHIQLTVHEPEOQOOLER---SKLSGSSSSS 2921
 QY 1371 TTDKNNENRESEK---KQORTSTQINGKDNKPKIYLGECLKEISESRVSVGNEPK 1426
 Db 2922 HADERHRREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 2959
 QY 1427 VNNINKIIPENDIKSLTVKSALRPPIINGDYIMEDFENRNSSETKSHLLSSSDAEGNYRD 1486
 Db 2960 VDDNSVDMDEAGALEAQ-----LMSDPFKTPISE----- 2990
 QY 1487 SLELPSTK---ESDSQTTTPSASCPESSVNVY-----EDMEIETSEYKAKVTSSPIT 1537
 Db 2991 --EATPSTATYSDMDVFRPSDN--EDNNSVDMTKOGVSEOEOKHSKDKKKRRSK 3047
 QY 1538 SEESNSLNDPIDENGLPIKNKENVNGESKRKYITEVTMTSTVAVESKTVIVEKGDK 1597
 Db 3048 EENQEKLL-----QOQRESLPNAVSTSSAPRPGKLTIVAVQASK 3088
 QY 1598 QT-----VVSSTENCASKTV-----TTTTTTLKLPSTSGSVDIISVKE 1640
 Db 3089 HADLOLDAKISSAPVCKPSPSLPCLIGDDDDALHTPKAKPTTPSSRGDGLPSPREK 3148
 QY 1641 KTV---VTTVYDLSL-----TTGTLVTS----- 1662
 Db 3149 RLISPIPKPTPIANSSTLSTQSAETPVSGTVISSSALATPTTPSSTAAGVSAAPGLDNP 3208

QY 1663 -----MTVSEKYST-----RDVKYLMKFSBPK 1665
 Db 3209 TSASOQKKKESFIPGFDGLDRISASAVGISAENSTSLDNIADDEPIPIVAPSPRA 3268
 QY 1686 TRSGTALPSTRKFTYKSTKK----- 3270
 Db 3269 TKPLDKLEESKSRVLTISOETESAVSALLGESFGTSTTDYSLDGMEDMSVNELETPTL 3328
 QY 1708 FVLPNDDLKTLARK-----GIRFVRYPNYNAKRALDIWYRSP-----RPTGIT 1753
 Db 3329 VIAPDEBALAKALETAGERPASTLEP---EMERREAEPRPDEALESPVEVL 3384
 QY 1754 WRVRL-QTVKSLAGVSLMLRLMLASLRWDMAAKVPPGGSTRTETSETITTEILIKR 1812
 Db 3385 DPELNAVQSLKHEMDM-----DIKADTPQSERDQIDT--DTEENPDE---AD 3429
 QY 1813 DVGPRGIRFEYCIRKTIICPIGVPETPKETPPROKGLRSSALRKRETPKQIGPVIET 1872
 Db 3430 SSGP-----SLKIDETVOSSSPKESISNNS-----PTPRETANIDI-- 3466
 QY 1873 WVAEELELMEIRAFARKEVEKKAQAVEQAKKRLQOKP-----TVIATSTSPST 1926
 Db 3467 -----PNVESQPKLSNESTPQSVYITKLPPLDTPKTVPAGLPSPVYKIE 3510
 QY 1927 TSTISPAOKVWAPISGSV---TTGTRMVLITKVGSPATVFOONNFOPTATWVKOG 1982
 Db 3511 PRTISKLOPLVQVQVTVLAPRHSTGSGISANSYINLDLNVISSCNTSAASATASASA 3570
 QY 1983 QSNQVVOVOQKVLGIIPSSSTGTSQOTFTSFQ--RATVPIIRPPTSOGGTTSN---SQV 2038
 Db 3571 SISGPSASQNAH---POASTPRKQGRITPOQARTOSLMOPTISIPEOTPRFAVPM 3627
 QY 2039 ITGPOI---RPG--MTVIRPLQOSTL---GKALITR-----PVMOPGAPQ 2078
 Db 3628 VLSQSHHPQOPGTVMGIRAPRPHSLRHSRGRVADSRVLGOLSPVGRFVQSP--SPQ 3666
 QY 2079 VMTQIIRGQVSTAVSAPNTVSTPQOKSLTSATSTNIOSSA---SOPRRPOGOVKLT 2135
 Db 3687 ---QVOQTQOQNALITSPQSNISPLASPTRVLLSSNSPTSKVNSYQPRNQ----- 3727
 QY 2136 MAOLITOLGNGGNOGLTVUIGOGOTTGOLILPOGVTVLPRGQOLMAQMANGTYQ 2195
 Db 3738 -----VPQRPKSAVAEVOITPROMTIT---PLOKMTPIOVNHP-- 3774
 QY 2196 FLTPRLATATATSTTTTSTVSTAAGTGEORQSKLSIPOMOYNOXKTLPRPAOSSVGRKA 2255
 Db 3775 -ITSKVTVVOPQATOSQVASSPLGS-----LPHKKNV---LMAON-----QO 3816
 QY 2256 QPOT-AOPSARPOROTPOSPAPORVOTQREVOQTUVSSHPSEADORTNAOSSKPOVA 2314
 Db 3817 QPVYAKMTAHNOQNOHQNOHQNOHQNOHQNOHQNOHQNOHQNOHQNOHQNOHQNOHQNOHQ 3862
 QY 2315 QSQPOSNVQOSPVKVPVQSPQTRIRPSTPQSLSPGQ--QSOVOTTTSPRIPTSLQIP 2373
 Db 3863 APQOHMOHQNOQOQOHNQOHNQOHNQOHNQOHNQOHNQOHNQOHNQOHNQOHNQOHNQOHN 3918
 QY 2374 SOGQPOSOPOVUS---STQTLSSGOTLNOYVSSPSRPOLOIOPOP--QVAVAPOLOOQ-- 2428
 Db 3919 -QHQQOQOQNOQOQOHLNLSOQ 3973
 QY 2429 -----VOVLSQISOVVAOIQAOQSGVPOQIKLOLPQIOQSSAVQTHQION 2476
 Db 3974 PQGVGHGSGSIFASQOHNQSLPAR-----GVPOQ--QHPOQLSHSSPCKKNTLVSV 4024
 QY 2477 -VTVOASVOEOLQRYOQLRDQOQKKQO-----QIEINV--NTP 2513
 Db 4025 NOGVQPRALITRVSQHPQOQ 4084
 QY 2514 SKLLIKVEIT-QKVY---VKNNAVIEHLQKQKMTGAKEBENQMLVCQVMKYIIDKI 2569
 Db 4085 -KIIVQOHIVAONOVPRQOTQGNALHYPONQOKSTPPGH----- 4123

QY 2570 DKEKQAKRRKRRESVQKRSKONATKLSAL-----LKKHKQLRAEILKKRALLDK 2622
DB 4124 -VEPTPMSAQKTSSESIVIRPTPTGLAVISANTVGSLLTEMLIKISOPKODELIEQ 4182
QY 2623 DLQLEVQ-----BELKRDILKK-----EKDLMQLAQATAVAPCPPTVPLPAPPAP 2670
DB 4183 DSK-EVDSDDYSAAEVENIDSYIKKIDLPPLASKDAKRAVEMQAI-----AP 4226
QY 2671 PPSPPPPVQHTGLSTPLPVASQKRRREKDDSSSKSKKKM 2715
DB 4227 APIPNQPGNQ--SMAQETALPTMSVNSNDHDTBEDTETROL 4269
RESULT 14
ABB66811
ID ABB66811 standard; Protein: 6815 AA.
XX ABB66811;
AC ABB66811;
XX 26-MAR-2002 (first entry)
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 27225.
DE Drosophila melanogaster polypeptide SEQ ID NO 27225.
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
PE 23-MAR-2001; 2000US-191637P.
PR 23-MAR-2001; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-6556860/75.
DR N-PSDB; ABL10914.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 27225; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116175) and the encoded proteins
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB85773-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 6815 AA.
QY Query Match 3.8%; Score 567; DB 22; Length 6815;
Best Local Similarity 17.1%; Pred. No. 2.6e-19;
Matches 609; Conservative 562; Mismatches 1238; Indels 1150; Gaps 145;
QY 2 VSEEEEDDGAETQDSEDEDEME--DDDDSDYEE----- 39
DB 1210 VTELQEPNQEVEYETRMS-DFYEKVLTELPSEMSDKPRTKRLHKEKGDEVQVLEIVESV 1268
QY 40 ---MEDDDDAASYCT-----ESSFRSHSYSTPGRKKRPRVHRPRSPILKEKDIPPLE 89

DB 1269 VAPVEEDLKMPALADVIALDEKIR-----KRVRAKDS---KOYEIEIE 1314
QY 90 FPKSSEDLVNEHNMVIAIYELRN-----GTVLRSPFREDCAALY 136
DB 1315 TEKRED---IPDEARVYITN-EVSGDTIDGPASTTEAPKRSKAKKEKEIEYVIA 1370
QY 137 SQEDCTLAEMHNVLLAKVLRREEDTSNTTFGPADLKDSVNSTLYFDGMTPEVLRVYCE 196
DB 1371 EEARLDHVSEIY-----EDYLRTPFRESSKEKI-----PSFTTYVE 1408
QY 197 SDEYHNHVLRYQ-EADYPTGPEVNTKY---LQFLVDQFLTNARELMSEGVYQYDD 252
DB 1409 -DEIVNPVLPKIRKTVDPVDRVPKDKKKKIDQKKIKIISFEPPTSEDSTIEEYTPKISE 1467
QY 253 HCRVCHKLGDLCCETGSAVYHLECVKPRPLEVEDEMECEVCA-----H 298
DB 1468 HD-----EDLOTDEYSDVYKDSLRSKSKSKTKKOKKE 1499
QY 299 KVPG-VTDCAVAEIQKNKPYIRHNEPIG---YDRSRKKYFLNRLLIEEDTENENKKIMY 354
DB 1500 SLPGPISLYTIRIETPTP---EPITEKITYEDGKEVYAVIKRRI-----KKKA-- 1544
QY 355 YSTVQVLAELIDCLDKDYWEALCKILEMREELHHRMDITEDLTNKAAGSNKGSFLAAN 414
DB 1545 -GPREYLEIYEIETEDNNPADV-----VIRIETTPSIDSKPOEDHKI----- 1587
QY 415 EEIESIRAKKGDIDNV-----KSPETEKKDN---ETENDSKAEEKRE----- 456
DB 1588 -QVQOEKKPKTESIDNTYQKILDOEIPQVDHKEKKAIVLETSPESKAKKKKKHKKTTTE 1646
QY 457 -----FEEDOSLEKSDDKTPD-----DPEQKSEEPTEVGDKNGSVAN 497
DB 1647 VIDGPIVIEVTIOETEDDEDFKPEDEVLTKIEDHEAEPAEQVFEITET-KAIDE 1705
QY 498 LGDNT-----NATSEESPEGRSPVGCSE-TPDSSNAEKVAVASEL-----PQ 542
DB 1706 LSEVTVAEITDEQPOEELVPAQOEKKPIKOKRKLPEDEVNTYVLELETPTEPTIPE 1765
QY 543 DVPEPNKTCSSNMTSATTTISIQPNLENSNSSELSNSESASAKAADPENGERE---SH 599
DB 1766 DADKPPQVIEDISENNQVQI---IEEDGT-----EKQVITKKKKSPKIGPEQVFEI 1817
QY 600 TPVSIQIEIVDFTSEKSTGELSPGAGGASSTRITRLRNPDSKLSQKSOQVAAA 659
DB 1818 TETPSPDEPLAEVITVELTEE---GLNKDI-----VIOEKKTVKPKPKLKPEDIOSY 1867
QY 660 AHEANKLFEKGGK-----EVLVYNSQGEISR-LSTKKEVIMK--GNINN 699
DB 1868 VIRVLEEFNEPQMPASTEKPIIEDIASIETVPTIEDGTKEVEYKVKKVSROGTKNQ 1927
QY 700 YFKLQEGKQYRVYHNOYSTNSFALNKQHREDHDKRRHLAKFCLTPAGEFKWNGSVHGS 759
DB 1928 VFEL-----IETKTSDEPLAEVITLELSDG-----KS 1994
QY 760 KULTI-----STRLTITOLENNI-----PSSFPHNMAHSHRAMIKAVQMC 801
DB 1955 QEVITLIREKKPIKTIKLPEDVESYVVVLEEFCEPQSEPEPEPREGAHEKTKT--KT 2012
QY 802 SKRPEFLALAILCEAVKPVMLP---IMREFIGHRLHRMSIEEEKKQVKKKKKK 857
DB 2013 KKRK-----KPLVKAPEENVILIEEAPFELVIEINIVEIGEYKQVYKTKKLK 2059
QY 858 EEEETMOQATWVKYTPPVKQVWVKQGEYRVTVGYGSMWISKVHYRFXPLGNTNVN 917
DB 2060 KKEPKREYLEIKETY----- 2075
QY 918 YRSLEGTKNMDENMDESKRCKSRPKIKIIEPDSEKQVKGSDAKGADQNMENDISK 977
DB 2076 -----BENKPEGD-----IEITTELVLPRESPPASD--DQPVIVQK 2110
QY 978 ITEKK-----DDVKKLLDSDSK-PCKE-EPMEVVDMDKKTSHSNVCOSSQVQVYVN-- 1027
DB 1027

Db 2111 IKKKPKVKKDDIKYIIOOLIEOITKTPLEEEPEPEMDSKKPKKVKVSHNKKITEVIDGL 2170
QY 1028 -----VSEGFHLRT-----SYKKTKSSKLDGLERRIKOF 1058
Db 2171 PVTHFNEVDIYSEPEMDMEPKTLDEIKETIQLPDDSSKYLVINISDFEBAKPIKOP 2230
QY 1059 TLEKQORLEK-----IKLEGGIKIGIKTSTNSSKLSSEPVITAKECQSDSM 1107
Db 2231 TODPIKKEPKLKKKKVVEYVSLAFDHTVKKVSEPTLEGTVEKVVKKKVRKRGSK 2290
QY 1108 RQ--EOSPNNNDOPEDLLIOCSOSDSSVYLMSPSHTKNLYKDRVL--DDV----- 1157
Db 2291 DHFEITETTSEDRPLEAVYVELVSEDLSEKPKHERKIVKPKQOLKDDVEEYIIN 2350
QY 1158 -----SIRSEPKCPKONSI-----END-----IEEKVSDLA 1184
Db 2351 IIEFIOPIPVGLVEDVEKQKKEKPKKSPITTYIATBOEDNDNNYDALVKEDLDOP 2410
QY 1185 SKGOEPTKSKTGNDFFIIDS-----KLASADDIGTLICKNKKPLIOESDTIV 1233
Db 2411 ERALEKSSPLEYTIYSEESVGEOKOPKPKKISKPKSI-----KOPESVDSKPDYLV 2463
QY 1234 S--SSKSLHSSVPKSTNDRAT-----PLSRAMDE-----GKLGDSNSGTLESS 1280
Db 2464 NVISEESIIDEPIEDYVTEBAEEKSEEPTEKVEELETAVEREYTDGGETTKQSV 2523
QY 1281 DTWSIO-----DSSEBDMI-----VONSESISOEFTTR----- 1309
Db 2524 TKRKIKKLVGPKKEIIEIVETKTGDTPEYIVYTEEVOEKSKAPEKKAKYKAKKI 2583
QY 1310 -----BODVLEPLKCELVSGESTGNCEDRLPVKGTANGKPKSO--OKKL 1354
Db 2584 PKDDLQDYIOKLIEODIPKLEKYEKIDLEBPVKMK--RKPICKVKVQSEOPKEETEPEPI 2642
QY 1355 EEPVNCSCDO-----IKLNTDJKNNKNERESKK 1385
Db 2643 EDKVEKEISEVSDSEPKITVAVKEFIREKPEKPEIYVLEETVESKREPEEGKVR 2702
QY 1386 GQRTSTFOJNGKDKPKIYKGECLKEI--SESRYVSGNVEP-----KV 1427
Db 2703 EKVYKTKTI--KONRGSEVVDIIVEIDPTNDSVITVTTVTEPPDOPGVKOKRT 2760
QY 1428 NNINKIIPENDIKSLYKESAIRPFGND--VIMEDFNERNSSETK-----SHLSSSD 1479
Db 2761 KKKKKDEVEFVKRNV--IEEBAPOGEGSDLVIEDFVFKPSEKRRKKKPIKDKH--TSVE 2817
QY 1480 AEGNYRDSLFTIPSTKE-----SDSTOTTPSPASCP--SNSVNOVEDMEITSEVKKYTS 1533
Db 2818 EETPHEDEVLIESVPEDSPUSDLTIVDSVP--IEEPEPNKVNQIEDTKPREKKKRPSP 2877
QY 1534 SPITSEE-----ESNLSNDFIDENGLPINKNE--NVNGESKRTKVIIT-- 1573
Db 2878 AKLIEENVPEOTVEKPLEALHTSDLEKRPVQEFISISIKEEOKHHTPEKKSSKISSEQ 2937
QY 1574 -----EVTMTSTVAESKTVIVKGDQOVVSSTEKAKSVTTTTVTYKLS 1624
Db 2938 PKOPSTOYEIYSVEHDLKEEKEKPTVOYIOSETNVEETKD--DTGVKHKQVTTKRMRLR 2996
QY 1625 PSNGSGVDIISV-----KEOSKVVVTTTVVDSLTTT 1655
Db 2997 PAGEGELEIIEVARDOPPEAEITIVVEPEVNODEKPREKKKTKTKKODIHDIYO-- 3054
QY 1656 GGLTVNTSMYVSKYSTDKYKLMKFSRPPKTRSG--TALPSYRKPVTKSTKSIFF--LPND 1713
Db 3055 --KLIELETKTELEKYEKLEIEFPIYKDKPLDSPIDVLDESFKVOKKDKKSSTKVPNE 3112
QY 1714 DL---KLLARKGGIR-----EVPYFNNAKPA--LDIMPY----- 1743
Db 3113 ETPVOEOLYAKVNVVEEBAPEQPEIPVOILEYKPVEDVVKEVITEDEKRPVOEKTTRVLK 3172
QY 1744 --PSPRPTFGITMKYRLQYTKSLAGVSLMLRLMLSLRMDMAKVPFGGSGSTTEISET 1801
Db 3173 IGPEOETFKIT-----MIESEBNDSSVTIV-----DEEPIASPOSIEEHPDOSKE 3219

QY 1802 EIT-----TTEIIRRDVGPYGRIFREYCIKKII--CP-----IGVPEP----- 1838
Db 3220 KLAEPKKTAVKVKKDDLSOI-----VKLIEEIPKVDLEKYEKEMEKVKLVTS 3272
QY 1839 --KETTPQOK--GLRSSALRPKRPETPKOTGPVLIETVVAEELWEIEMAEV 1892
Db 3273 DSIEEPEKRPDKSOPISVLPDITTKPKKTKTPK-----TPKE 3308
QY 1893 KEKQAVBQAKKRLLEOK--TIVATSTSPSSSTSTISNAQVNVAPISGVTTGK 1951
Db 3309 DTDOQVDEPETTVVDTTIDELPOTQOPEDEATQITSAOEKSTODDTKOT 3368
QY 1952 V--LTKVSGPATVFOONKNEHOTFATWVQOGSNSGVVOOQ--KVLGI--IISSTGTS 2006
Db 3369 VKHKTKNDOTKSVTSELSLPEVHKDYOISI--IHEELVEEOPKLEKLEVRIDEVAEV 3424
QY 2007 QOTFTSFQPTATVYTIERNISGSGTTSNSOVITGPQIRPGMTV-----ITPPLQ 2056
Db 3425 EES--OPIVEEV--EDEEPOPATEETVEDVTKPKSKKKVKKTKTDHDELIKKLE 3477
QY 2057 QSTIGKAIIRTPVWVQCAPQOVMTOIIRGQVPS-----TAVSAPNTVST 2102
Db 3478 QE--IEKTELEKYEKIEFDVPKKLPEFALPEIKIERKQKPYVTIADTVKRYALK 3536
QY 2103 PGOKSLTSATS-----TSNIO----- 2118
Db 3537 PSKKREKAEELVQOLPKFRUKARWVLEYPPAPLIPKPTDIGAIKONGELSRNIEEAE 3596
QY 2119 -----SSASQP-----PRPQGO--VK 2133
Db 3597 ILKFRPHKTKIKKIKDKLEKVELEKYEKYISSEEPPEKTPYKPKKAPKPEKQOEYVK 3656
QY 2134 LTMQOLQLTGHHGNGSLVVIQOGOTGOLQILPGVNV--LPGGOOLMAAMPNG 2191
Db 3657 LKL-----GKGKKKPEEAPRNVTLKNIPKQPOEV----- 3687
QY 2192 TVOREFLPLATATATATATTTTIVSTTAAGTGEOROSKLSL--OMQVHODKTLPPAOSGSV 2250
Db 3688 -----EEVELKQKKEVEIYEEQKKRKGDFV 3716
QY 2251 GPKAQPOTAOPASRPOPOTOPOSAPDEVOTQPEVOTOTTSSHVSEAOPTHAQSKP 2310
Db 3717 -----VEPEEPESEFDEVPDELEQIEHEIPEKVKPKSKTYKKPKDKSSEPTIVS 3770
QY 2311 QVAA-----OSOPSNVQOGSPVR--YOSPQTRIR--PSTQSLSPGQSOVOYTTTS 2359
Db 3771 EIVAGVPEKEEALPEODVKFKRPERDAPETDSEIKLRPVQASKDENPDQALVTPKAE 3830
QY 2360 QPIPIQ-----PHTSLQIPSOQPOQOSPOVOSTOTLSSGOTLNOVSVSSSRP 2408
Db 3831 EPIQOELDKAIDEXKPKKSK--PKVQOPKEQOLAEPEPEFVSVKEEALVDKP-- 3885
QY 2409 QLOIQOPQOYIAVPOLQOQOVLSQIOSOVAAQIOAQQSGVPQOIKLQPIQ--IQOS 2465
Db 3886 --IEIKPKDVKKVEKKPKPEAPVSEV--VIEEPEKPEVPEIPEYKITTIVLEPE 3939
QY 2466 SAVOTHOIONVYVOASVQOQOLRVQOLRQOQOKKQO--QIINNTPSKLLIKYEII 2523
Db 3940 DAPREHOYKVIDFEROETTEEVEIIEEKVVRKKPKQOPEFEVTLKEPEKEQIOPDV 3999
QY 2524 QOQVMMKNNAVIEHLKQKSM-----TPAREENQ-----RMIVCQNOVMKYLIDKI 2569
Db 4000 SAEISLP--IEEPBOKPREYVELKITOTTPPEPNDVQJAVAEKKTVPYKVKREDKI 4055
QY 2570 ---DKEEKQAARKKRKEEVEQORSKQMATKISALLFKHKEQOLRAEILKRALLDKLO 2625
Db 4056 VVVAEERKQOPEVEITIVE--VEKQEEKKSKSEKPSYEFKISSETOSIE-----EKP 4103
QY 2626 IEVQBELKRLKIKKEDLQALQATAVAAFCPPVTYVLRAPRPPSPPPPGVQHTGL 2665
Db 4104 IEVAEAPPEETPKVYKVAEKFDYSFTLKETDEEKVITVDDQPEE----- 4150

```

QY 2686 LSPPTLPVSAOKRRREENDSSKKKKMISTSKET---KKDTLXCTCKTPVDESKF 2742
Db 4151 -EAPVAVVTKRKRKEEPAVEAEVMTPEKIVETSVETAIKOKTK- ----KPKKDEEA 4204
QY 2743 YTGDCRCQMWYHRCVGILOSFAELIDEVY--CPQCSYEDAMTVTLPTLEKDEGLKRY 2800
Db 4205 QL-----AIKVESEAPVAEEVSEAPESKIVEEVL- ---EKKPEETIV 4248
QY 2801 LNSLAKHKWAPPELEVDN-----DADYGVYKE--PMDLATMEERY 2842
Db 4249 SESEP-----KEEPSVEQFTVKRRKPSVTFADEPATEITIVIKESKPAEVTEDAH 4299
QY 2843 Q-RRYEEKLTFEYADMTKI 2860
Db 4300 KTKKPKKKYTDVAEAEELKI 4318

RESULT 15
ABB67502
ID ABB67502 standard; Protein; 3257 AA.
XX
XX ABB67502;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 29298.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL11605.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 29298; 21pp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABB57737-ABB872072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 3257 AA;

QY 3 SEEEEEGDAEETODSEDEDEKEED-----DDSDYFE---E 39

```

Db	238	S0S5TEYBDEYGGDD5--DEEDPEVVIDSTDTSKTTSVADTVTPPVVSDDEGE0QSELN	293
Qy	40	MEDDDDDASCYCESSEFSRSHSTYSYTPGRKRPVHPRSP1-----	79
Db	296	QJRHAND0SP5LST5TYERL0SYLNSLSA5PPIYK0K0SAKKSNSNEKKPSS0QVHAN	355
Qy	80	-----LEEDDIPPLEPKS5EDLWVNEHIMVAVIYELVRFNGTVLRLSPRF	122
Db	356	EQSAASDEANKREKDLBP--PEASSAKSSAKSASGSRKLYLVANN-----	400
Qy	129	EDFCALV5OECOT-ME--MHVVLKAVLEEDTSNTYFG-----PADLK5-----V	175
Db	401	-----ATLEAVEM5IDLAD0IMEIDVER5YEFERKSASKSLSRSKIPAD5SISOLGHV	456
Qy	176	NSTLYFIDGMVPEVLYRYCESDEY-----	201
Db	457	RKLTYSSERLQPELIGLQLRKRSKE5LPKGRPRGOKORKNRGATYBESKANCROGE	516
Qy	202	-----HHVPOAEADPYGVVEKIKVLOFLVQ0LFTN-----IA	238
Db	517	PKYKRRGRPKRIIPREKAKT--AETENTISLNTWNPPL5DTSKENPETYVLEVPK	573
Qy	239	REELM5E-----GVLOYDHCRVCKRLDGLLCCETCSAVYHLECVKPLBEVPE	287
Db	574	ODELVSDDLNAKELTGSNLLPDDDTIEMASNHQEDTLKCAPRVALDKSEST--PKVE--	629
Qy	268	DEM0CEVCAVKVPCVTCVCAVLOKNKYIINHPYIGVDRSKRKWFLNRLIIE0TENE	347
Db	630	-EOLCKVTPSPDUALDEKSVESAKN--HILE--DKDKD-----EET0KE	671
Qy	348	-----NEKKIMVYSTKVOALAEIDLCLDKDYKAEFLCLIEEMREELIHRHMDIEDL	398
Db	672	SPNGSKETNENSV-IYVNEVELP-----AKAEKAKAGNIVEES0QJAEDEPKAEI	725
Qy	399	TINKANGSKSF5LAANE5IL5IRAKKD-IDNVK5PELETCKDNFENDSKDAEKREE	457
Db	726	LALEVKGK-----VEANEVSAT5VOGE0NPVEIYKELEQ-----ETISEVPRAND0SS	775
Qy	458	FED0SL--EKDSDDKTPDDDOPE0GK5EETPEVGDKGNSVANGLDNTNATSEETPSE	514
Db	776	VEDQTLAKENPVKEPR5PVKAPSSK0DRP-----AEENLPAP	813
Qy	515	GR5PVGCLSETPDSSNMAEKVVASLPO-----DYPEE--PNKTCESNTSATTT5I	564
Db	814	D0DPIE--0OKTPVAKNO0HDEHNEAPKAE5LVS5DIPSS5VTP5PK--RHH5SPANTPK	871
Qy	565	QPNLNSNSSELN50ESAKADDPENNGEREHTVPSIOE--IYGDFTSEK5TGL5	622
Db	872	SKETI5AL05VPRRLAKRDKA-----TQON-LRERSKRTKLTELTLTMDTMRSSPRLG	926
Qy	623	ESPGAGK5ASG5TRITRLRNPD5KLSQLK5Q0VAAAHAANKLEKGEVLVNS0GET	682
Db	927	R5P-----ASH5SHERS5PEKKVTV5KLAKLD	954
Qy	663	5RL5TKKKEVIMKGNINNFKJG0G5KRVYINQ5T5N5FALNKHQHNEDDKRHLAKHF	742
Db	955	ITIDKEKIELEKS-----	967
Qy	743	CLTPAGEKKNMG5VHG5KVL5TL5RLTLIOLENNIP5FEFHPMASHRANWIKAVOMCS	802
Db	968	-LPPASERK--DVKITKTTA5T5T5--LIDEN5S--SKTEMKMLK0KRPPLAKKM-S	101
Qy	803	KPREPALALALLEC--AVKPVVMLPIKREFLGHTRLHRMT5IEREEKYVKKKEK0EE	860
Db	1018	RT5E5EVKALAD5NEDIP5FISIKVCEHNL-----TSSE0QD-----EKE	106
Qy	861	ETM0QATVAVKTFPVKQVWK0KEEYRVYGG5MSI5KTHYRPRKLP5GNTNNYK	920
Db	1061	ELLCPK0ID0T-----NTDLE0ST	108
Qy	921	SLEGTKNMDENNDESDRKRC5R--5PKRIKIEP5EKDEVK5GDAKAGADONEMD5IKT	979

```

Db 1081 ALETDEQVEK--RSNRKSRIRNEKFTENDTILSDHLD---AKKAENASLEISM-- 1132
Qy 980 EKKDQVKEILDSDDKP--CKEEMVED---DMKTESHVNCQESSOVYVNVSEGFHL 1034
Db 1133 -----RKCOTLETOQSDPYTAKNRKNSGRLSRKES--VINAAS-- 1170
Qy 1035 RSYKAKKTSKSLDGLERIKOFTLEKORLEK-----IKLEGGIKGIGKTSTSSKN 1088
Db 1171 ----EKDKSPSALSOSTER--KOLNENPSKKDKTQESGNKKEAVVGPJDKTETSSTN 1224
Qy 1089 LSESPLYTAKKEGQSDSMRQEOSPNANDQPELDLQGGCSQSVLRMSDPHTTNKLY 1148
Db 1225 IIDK-----KSNEFSDAM-----QPSDRL---NOKESATFKUSISSPKKIM 1265
Qy 1149 PKDRLVDVS---IRSPETKCPKONSIENDIEKESVDLASRGOEPTKSKTGNDFFIDS 1205
Db 1266 DQDKDLALSKGDSNPTIRDTGEDSRQTDKKNQEND--TKHEEDSSKUKAN---IDET 1320
Qy 1206 KLASADDIGTLCKKKPLIOESDITVSSSKALSHSVKSTNDR-----DATPLS 1257
Db 1321 KSSSE-----KDAEPIKSSSQ---DSAKPRL--SKPKSRKKKKKKKKKPNDSIAES 1367
Qy 1258 RAMDFEGKIGCOSE-----SNSTLENSDPTVSIODSSEEDMIVONSNEISQFRTREQ 1311
Db 1368 ---DIEGGOVNTETVQATCSTPESNKKDMVKSDETNEE---PMLSEIGIRKRGQ 1420
Qy 1312 DVEVLEPLKCELVSGESTGNCEDRLPVKGTENGKKRPSOOKLEERPVKNCSDQIKLNT 1371
Db 1421 AFHIEHPKDDHITTPQN-----ENOSIAGVNEFKOYPLPESVESDIPRIMKIPIKTYLMCT 1475
Qy 1372 TDKKNNENRESEKGGORTSTFOJNGDKNPKYIKGECLEKIEISRVASGNVEPKVNNIN 1431
Db 1476 KNTKS-----LLASDEPDILYEPQ--KLITTSK---GDSNDPDLNAN 1513
Qy 1432 KIIPENDIKSLYKESAIRPFIINGDIYMEDFNERNSEFTKSHLSSDAGNRDSEJTL 1491
Db 1514 NL-----ETSTQDPKEHEFS---DOTTDSNIT 1540
Qy 1492 PS-----TKESDSTQTT--TPSASCPE-----NSVNOVED--ME 1522
Db 1541 PSCTKKSQIVFTPTPKSSQOTKNSFITTPNRS--PKSKRNVSKAARLNDNFEESSQNASE 1599
Qy 1523 IETSEVKKATYSSPTITSEESNLSNDTIDENGLPINKNENV---NGESKRTVIT--- 1573
Db 1600 SSASKYQKELRPTASCRLRVLKRTPTSSLPNSRSKSIFFKTPAKSKRLTLKLESMEK 1659
Qy 1574 -----EYTMSTVATESKTVIKVEKDKQTVVSTENCACKST-VTTTTTYTK 1621
Db 1660 TTSREPSVSLGEVNPDSPPVAESVAVLHESDRDLESNEIPEEVEDETEEASAEJTDNK 1719
Qy 1622 LSTPTGGSVDLITSVKEOSKTVVTVTVTDSLTGTLVTSMTVSKEYSTRQV--KLAK 1679
Db 1720 LKKKEHDHELEVNDICAANKNPT---DDSTKASSNKSTDSVLOE--TKDELSNLSLN 1774
Qy 1680 FSR-----PKTNSGIALPSTYKRFVTKSKSIFVLPNDIDKUL--ARKGIREVYNYNA 1734
Db 1775 ATOGEDPTIKELTEEVVNNKTVEDESKROEI-----LKLEDPNAALEEDTASTAKA 1827
Qy 1735 KPALDITWYPRPRTGCIWRYRLQTVKSLAGYSLMLRLMLASLRMDDMAAKVP--FGGS 1793
Db 1828 AEMMDLY-----IKESINXSVLAEPETDVTDEELAOQPIPNSSSE 1868
Qy 1794 TTEETSEITTTTELKRRDVPYGIREFEJCIRKICPIGVPETPKTEPTPOKGRSSA 1853
Db 1869 TTSYVDNDPEPSTSSVVK-----SLRK-----READSSQD---EAA 1902
Qy 1854 LRPKRPETPKQ--TGPVITETWAEELLEWEIRAFAEVYEK--KAQAVEQOAKKRLQEQ 1910
Db 1903 KKKORQDVEKSLTG-----KKEQVKPARRROLAEEBERPLKRSKTESSEAKSTVOGK 1954
Qy 1911 KPEVLA-----TSTSPTSST---TSTISPAQKVAVAPISGSVTTGTKNVLTJK---VG 1958
Db 1955 YSTIGNETIMSTTAPIRETNREASTSPSARK-----SAQBAKHVETTKHITIG 2006

```

```

Qy 1959 SPATVTFQONKNFHQFATVWVWOGQNSGVVOYQOQVGLIIPSSGTSGOQFT----- 2011
Db 2007 PPG-----KKLH-----SDSPAEEVKKPMVOTJLSTLSLQKRPSTLDDGSP 2049
Qy 2012 -----SQPRATVYTIPIRNT---SGSGGTTNSQVYITGPQIRPGMTVYIRTLQOSTLGKAI 2064
Db 2050 KIRKSLKKSIAIDENIDGDOSIFSSSVLKNNTSVVAPRKYNISVLSQSKDTQVETPAAS 2109
Qy 2065 IRTPVWVQCAPQOVMTQ-----IIRGQPVST-----AVSAPNTVSS 2101
Db 2110 SETPIILTK---KEKLTQKSTKKEPKKTESKKSLSLVQSPQMKTKQSEAVSGPKLNL- 2165
Qy 2102 TPQOKLTSATSTNIQSSASOPRPOGOQVKLMAQLTQLTQGHGNOGLTVVI---Q 2157
Db 2166 ---KYLKSETESSR--KTVSTVGRKQIGQL-----EVLKPPSRKSEELVEAISKKQ 2215
Qy 2158 GQGQTTGQLOLIFQGYTVLDPG-----GQOLQOAMPNGVYQRELFPLATYATTAATT 2213
Db 2216 SQVQRLSKIDGRKSEGTSLPQDPVSKSETALKALPK-----ETEFVYODA 2261
Qy 2214 TVSTTAAGTGEORQSKLSPQMVOYHODKTLPPAQSSVGPAKAQOPATAQSPAROPOTQ-- 2271
Db 2262 EIKMSKGRGHQ-----AVKNTKTEOP-----KSKPKTEVRSLOABATELM 2304
Qy 2272 ---POSAPAEVQOTPEVOTQTTVSSHVPSEAQPTHAOSSKPOVAASOPQSVNOGOSP 2327
Db 2305 DMSDQSDVSDIATIPFESGIFNVPGHM-----TFAISNNSIATPTPRMDSQ----- 2354
Qy 2328 VVQSPSQTRIRPSTPSQSPGQOSOVQTTSQPIPIQPTSLQIOPSOGQPOSOPOVOSS 2387
Db 2355 ---RMAKSERFTT-----VSDQKKPIRESQTLKRRARGNRGRNDPLYSKR 2395
Qy 2388 TQTLSSGQTLQNVQSVSSPSRQLOIQOPQOYLAVPQLQOQVOYLSIQSQVAAQIQAOQ 2447
Db 2396 ---KAGEAEDGTAIVNPKPRREDEEDHPQ-----QN 2424
Qy 2448 SGVPOQIKLOLPIQIOSSAVQTHQIONVTVQAAVQEOLOQ-----RVQO-----LR 2495
Db 2425 DHVQESAFAFAPVKITPAASSV---IPQVVRSGVNTVPQNIISPRKLCVKLNRPRYNKWL 2480
Qy 2496 DQOQKKQOQIEINVTPSKUL-----IKVEITQOVYMKHNAVIEHLKOKKS 2543
Db 2481 STQERNEDQGSRNV--TSLPLGSETSETDSAASMSSESILOQV--OSEPAIOPL----- 2532
Qy 2544 MTPAEREENORMVQNVQVMYIILDKIDKEKQOAKKKRRESEV-----EQKRSQNA 2595
Db 2533 ---PASQDSCCTL-----QASDLKIRESSAQOLAPIAAYDSPAADSS 2571
Qy 2596 TKLSALLFKHKEQLRAEILKRALLDKLOIEVQOELKRPDLKIKKEDMLQAOATAVAA 2655
Db 2572 TSPALDIAIPASQATAKATLNTALCPSTEKHLRDEPRLBESSKVAEBQKQOTQOAKCLPV 2631
Qy 2656 PCPPV-----TPVLPAP--PAPPPSPPPPEVQHTG--LSTPLLPVVASQKKR 2700
Db 2632 PIPEVASEPEDIHDESPNEMPMVAAPATPQPHAITEDAGDGTQIOVNTLGVSTSRPL 2691
Qy 2701 EEEK---DSSSKSKKKKMIISTSKETKKDTLYCICTPYDESKFYIGQDCP 2749
Db 2692 ELHSIPASDPGNPNAIGQTKMYSFLYPRY---KQSIDD---VGLDQC 2735

```

Search completed: November 20, 2002, 16:25:02
 Job time : 130.106 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:05 ; Search time 70.894 Seconds

(without alignments)
5227.100 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MVSEEEEDGDAAETDSE.....KIKGFKASRSHNNKIQSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002:*

1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*

2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*

3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:*

4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*

5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*

6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*

7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*

8: /SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:*

9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*

10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*

11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*

12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*

13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*

14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*

15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*

16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*

17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*

18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*

19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*

20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*

21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*

22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*

23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14333	100.0	2781	21 AAY57453	Human transcriptio
2	14170	98.9	2907	21 AAY57452	Human transcriptio
3	2695.5	18.8	557	22 AAB94078	Human protein sequ
4	1537.5	10.9	976	22 ABB68735	Drosophila melanog
5	1474	10.3	328	21 AAB54364	Human pancreatic c
6	982.5	6.9	238	20 AAY043023	Human secreted pro
7	981	6.8	237	22 AAU19446	Human diagnostic a
8	789	5.5	149	23 AAU78845	SCAN/KRAB protein
9	781	5.4	246	22 AAU19444	Novel human secret
10	672.5	4.7	997	22 ABB62887	Drosophila melanog

11	572.5	4.0	2768	22 ABB68397	Drosophila melanog
12	560	3.9	5533	22 ABB65772	Drosophila melanog
13	560	3.9	5560	22 ABB71160	Drosophila melanog
14	549.5	3.8	6815	22 ABB66811	Drosophila melanog
15	534	3.7	3080	22 ABB64877	Drosophila melanog
16	529.5	3.7	3257	22 ABB67502	Drosophila melanog
17	508.5	3.5	2897	22 ABB58514	Drosophila melanog
18	441.5	3.1	4498	22 ABB58595	Drosophila melanog
19	434.5	3.0	2977	22 ABB69480	Drosophila melanog
20	432.5	3.0	1878	22 AAM40239	Human polypeptide
21	432.5	3.0	2137	23 ABB39618	Human polypeptide
22	432.5	3.0	3266	21 ABB42491	Human ORF2255
23	430	3.0	1795	22 ABB6806	Drosophila melanog
24	430	3.0	2586	22 ABB6878	Drosophila melanog
25	429	3.0	3726	22 ABB63947	Drosophila melanog
26	428	3.0	1727	22 ABB95554	Human protein sequ
27	423	3.0	2285	22 ABB63057	Drosophila melanog
28	418	2.9	4019	22 ABE13839	Human lung tumour
29	416.5	2.9	2665	22 ABB28314	Human peptide #95
30	416.5	2.9	2665	22 ABB33490	Peptide #96 encod
31	416.5	2.9	2665	22 ABB18950	Protein #94 encod
32	416.5	2.9	2665	22 AAM54270	Human brain expres
33	416.5	2.9	2665	22 AAM6665	Human bone marrow
34	416.5	2.9	2665	22 AAM14533	Peptide #97 encod
35	416.5	2.9	2665	22 AAM26950	Peptide #98 encod
36	416.5	2.9	2665	22 AAM02259	Peptide #94 encod
37	416.5	2.9	2665	23 ABB36319	Human peptide enco
38	410.5	2.9	1878	19 AAM81170	Human Bat2-alpha p
39	410	2.9	3596	23 ABB40235	Staphylococcus epi
40	409.5	2.9	2951	22 ABB60291	Drosophila melanog
41	404.5	2.8	3256	21 AAY50976	Human cell cycle p
42	403.5	2.8	1798	22 ABB71695	Drosophila melanog
43	402.5	2.8	1589	22 AAM42025	Human polypeptide
44	401	2.8	2344	22 AAU37120	Staphylococcus aur
45	400	2.8	2246	22 ABB65319	Drosophila melanog

ALIGNMENTS

RESULT 1	
AA57453	
ID AAY57453 standard; Protein; 2781 AA.	
XX	
AC AAY57453;	
XX	
DT 22-FEB-2000 (first entry)	
XX	
DE Human transcriptional regulatory factor SEQ ID NO:10.	
XX	
KW Human; transcriptional regulatory factor; TCOAL; BLAST detection;	
KW bromo-domain; cell proliferation; cancer.	
XX	
OS Homo sapiens.	
XX	
PN W09957143-A1.	
XX	
PD 11-NOV-1999.	
XX	
PF 30-APR-1999; 99WO-TP02340.	
XX	
PR 30-APR-1998; 98JP-0137631.	
XX	
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.	
XX	
PI Jones MH;	
XX	
DR WPI: 2000-052940/04.	
DR N-PSDB; AAZ39033.	
XX	
PT Transcriptional regulatory factor containing a bromo domain and gene	
PT TCOAL encoding it	
XX	

PS Claim 1; Page 139-151; 154pp: Japanese.

XX The present sequence represents a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulatory mechanism. It binds to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which are potentially useful for the treatment and prevention of cancer and other cell proliferation disorders.

XX Sequence 2781 AA:

Query Match 100.0%; Score 14333; DB 21; Length 2781;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSSEEEEDGDAEETODSEDEDEDEDDDDSDYPEEMEDDDDASTCTESSFSRSHST 60
DB 1 MVSSEEEEDGDAEETODSEDEDEDEDDDDSDYPEEMEDDDDASTCTESSFSRSHST 60
QY 61 YSSTPGRRKRVRHRRSPILIEEKDIPLEFPKSSSEDLAVNEHIMVIAIYEVLRNGTV 120
DB 61 YSSTPGRRKRVRHRRSPILIEEKDIPLEFPKSSSEDLAVNEHIMVIAIYEVLRNGTV 120
QY 121 LRLSPFFEDFCALVSOEQCTLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
DB 121 LRLSPFFEDFCALVSOEQCTLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
QY 181 FIDGMTPEVLRYCESDKEYHNVLRYOEAADIPYGVENKIKVLOPLVDFLTNTNARE 240
DB 181 FIDGMTPEVLRYCESDKEYHNVLRYOEAADIPYGVENKIKVLOPLVDFLTNTNARE 240
QY 241 ELMSEGVIOYDDHGRVCHKIGDILCCETSAVYHLECVKPPLEVPDEMOCCEVCANHY 300
DB 241 ELMSEGVIOYDDHGRVCHKIGDILCCETSAVYHLECVKPPLEVPDEMOCCEVCANHY 300
QY 301 PGVTDVCAEIOKNKPYIRHEPIGYDSRRKRYWFLNRLIIEEDTENENKKIMYYSTKVQ 360
DB 301 PGVTDVCAEIOKNKPYIRHEPIGYDSRRKRYWFLNRLIIEEDTENENKKIMYYSTKVQ 360
QY 361 LAELIDLDLDYWAELCKLLEEMREIHRHMDITTEDLTAKARSNKSFLAANEEILLES 420
DB 361 LAELIDLDLDYWAELCKLLEEMREIHRHMDITTEDLTAKARSNKSFLAANEEILLES 420
QY 421 IRAKGGIDHWKSPETEENKNETENDSKDAEKREFEFQSLKDDDKTPDDPDEQCK 480
DB 421 IRAKGGIDHWKSPETEENKNETENDSKDAEKREFEFQSLKDDDKTPDDPDEQCK 480
QY 481 SEVGDFSEKSNGLSESPGAGKAGSGSTRITRRLNPDSKLSQLKSQVAAAAHAEANKL 540
DB 481 SEVGDFSEKSNGLSESPGAGKAGSGSTRITRRLNPDSKLSQLKSQVAAAAHAEANKL 540
QY 541 FKEKEEVLVNSOGEISRLSTKKEVIMKGINNNYFKLGOGSKRYVYNNQYSTNSFALNKH 600
DB 541 FKEKEEVLVNSOGEISRLSTKKEVIMKGINNNYFKLGOGSKRYVYNNQYSTNSFALNKH 600
QY 601 OHREDHDKRRHLAKFCLTIPAGEFEKANGSVHSGKVLITSLRLITITOLENNISSFLHPN 660
DB 601 OHREDHDKRRHLAKFCLTIPAGEFEKANGSVHSGKVLITSLRLITITOLENNISSFLHPN 660
QY 661 WASHRAWMIKAVOMCSRPREFALALILECAVKKPVVMLPIWREFLGHTRLHRMTSIEREE 720
DB 661 WASHRAWMIKAVOMCSRPREFALALILECAVKKPVVMLPIWREFLGHTRLHRMTSIEREE 720
QY 721 KEKVKKKKEKKOEBEETMOQATWVYKTPPVKHQWVKOGGEYRVLYGSGWMSISTHYRRF 780
DB 721 KEKVKKKKEKKOEBEETMOQATWVYKTPPVKHQWVKOGGEYRVLYGSGWMSISTHYRRF 780
QY 781 VPKLPGMTNNYRKSLGCKNNMDNDESDKRCSSRPKKIKTEPSEKDEVGSDAAK 840
DB 781 VPKLPGMTNNYRKSLGCKNNMDNDESDKRCSSRPKKIKTEPSEKDEVGSDAAK 840
QY 841 GADONEMDISKITEKKDQVKELLSDSDKPCKEEPMEDDDMKTESHVNCQESSQVYV 900

DB 841 GADONEMDISKITEKKDQVKELLSDSDKPCKEEPMEDDDMKTESHVNCQESSQVYV 900
QY 901 NVSEGFRLKTSYKKTYSKSLDGLERRIKOPLLEKQRLKIKIGCIKIGIKTSTNSS 960
DB 901 NVSEGFRLKTSYKKTYSKSLDGLERRIKOPLLEKQRLKIKIGCIKIGIKTSTNSS 960
QY 961 KNISESPVITKAKEGCOSDMROBOSPANNNDPREDLIQCSOSDSSVLRMSDPSHYTNK 1020
DB 961 KNISESPVITKAKEGCOSDMROBOSPANNNDPREDLIQCSOSDSSVLRMSDPSHYTNK 1020
QY 1021 LYPKRDVLDVSIKSPETKCPKONSIENDIEEKVSDIASGQEPYKSKTGNDFITDCK 1080
DB 1021 LYPKRDVLDVSIKSPETKCPKONSIENDIEEKVSDIASGQEPYKSKTGNDFITDCK 1080
QY 1081 LASADNIGTLICNKKRLIOEESDPTVSSSKSALHSSVPRSTNDROATPLSRAMDEGKL 1140
DB 1081 LASADNIGTLICNKKRLIOEESDPTVSSSKSALHSSVPRSTNDROATPLSRAMDEGKL 1140
QY 1141 GCDSESNSTLENSSDTVSIOSSSEDMIVONSNESIIEQFRTREODEVLEPLKCELVSG 1200
DB 1141 GCDSESNSTLENSSDTVSIOSSSEDMIVONSNESIIEQFRTREODEVLEPLKCELVSG 1200
QY 1201 ESTGNCEDRLPVKGTANGKRPQOQKKLEBRPVNKCSDQIKLNTDCKNNENRESEKKG 1260
DB 1201 ESTGNCEDRLPVKGTANGKRPQOQKKLEBRPVNKCSDQIKLNTDCKNNENRESEKKG 1260
QY 1261 ORTSTPOINKDKPKTYLGECLKETSSESRVYSGVNEPVNNINKIIPENDIKSLTYKE 1320
DB 1261 ORTSTPOINKDKPKTYLGECLKETSSESRVYSGVNEPVNNINKIIPENDIKSLTYKE 1320
QY 1321 SAIRPFINGDVIMEDFERNRSSETKSHLSSSDAEGNYRDSLETLPTSKESDSQTTPPS 1380
DB 1321 SAIRPFINGDVIMEDFERNRSSETKSHLSSSDAEGNYRDSLETLPTSKESDSQTTPPS 1380
QY 1321 SAIRPFINGDVIMEDFERNRSSETKSHLSSSDAEGNYRDSLETLPTSKESDSQTTPPS 1380
DB 1321 SAIRPFINGDVIMEDFERNRSSETKSHLSSSDAEGNYRDSLETLPTSKESDSQTTPPS 1380
QY 1381 ASCPESNSVNOVEDMEIETSEVKVYSSPTTSEBSNLSNDFIDENGLPINKENNNGES 1440
DB 1381 ASCPESNSVNOVEDMEIETSEVKVYSSPTTSEBSNLSNDFIDENGLPINKENNNGES 1440
QY 1441 KKRKVTIEVTMTSTVATESKTYIKVKGDKQVYSSSTENCAKSTVTTTTYTKLSTPS 1500
DB 1441 KKRKVTIEVTMTSTVATESKTYIKVKGDKQVYSSSTENCAKSTVTTTTYTKLSTPS 1500
QY 1501 TGGSVDIISVKEOSKTVVTTVTDLSLTGTLVTSMTVSKSEYSTRDYKLMKFSRPKKT 1560
DB 1501 TGGSVDIISVKEOSKTVVTTVTDLSLTGTLVTSMTVSKSEYSTRDYKLMKFSRPKKT 1560
QY 1561 RSGTALPSTYKRFYKSTKKSIFVLPNDLKLARKGIREVPYFNNAKPAALDIWPPSP 1620
DB 1561 RSGTALPSTYKRFYKSTKKSIFVLPNDLKLARKGIREVPYFNNAKPAALDIWPPSP 1620
QY 1621 RPTFGITWRYRLQTVKSLAGVSLMLRLIMASLRMDMAAVPPGGSGSTRJETSETETTT 1680
DB 1621 RPTFGITWRYRLQTVKSLAGVSLMLRLIMASLRMDMAAVPPGGSGSTRJETSETETTT 1680
QY 1681 EIIRKRDVGYGIRFEYCIRKIIICPIGPVETPKETPTPQKGLRSSALRPKRETPKQNG 1740
DB 1681 EIIRKRDVGYGIRFEYCIRKIIICPIGPVETPKETPTPQKGLRSSALRPKRETPKQNG 1740
QY 1741 PVIIETVAAEELLMIRFAERVEKEKAQAEQAKKLEQOKPVIVATSTTSPSSST 1800
DB 1741 PVIIETVAAEELLMIRFAERVEKEKAQAEQAKKLEQOKPVIVATSTTSPSSST 1800
QY 1801 TSTISPAQKVMVAPISGSVTTGKMYLTIVGSPATYFOONKNFOTFTAWKOGOSNS 1860
DB 1801 TSTISPAQKVMVAPISGSVTTGKMYLTIVGSPATYFOONKNFOTFTAWKOGOSNS 1860
QY 1861 GVVQVQKVLGIIIPSSGTSGOFTSFQPRATVYIRPNTSGSGGTTSNSQVITGPQIRP 1920
DB 1861 GVVQVQKVLGIIIPSSGTSGOFTSFQPRATVYIRPNTSGSGGTTSNSQVITGPQIRP 1920
QY 1921 GMTVIRTPLOOSTLGAIIIRTPVWVOPGAPQOQVWTOIIRQOPVSTAVSAPNTVYSSIPGCK 1980
DB 1921 GMTVIRTPLOOSTLGAIIIRTPVWVOPGAPQOQVWTOIIRQOPVSTAVSAPNTVYSSIPGCK 1980

Db 1921 GMTVITPILQOSTLGKAILRTFVMPVPGAPQOVMTQIIKGPVSTAVSAPNTVSTPQOK 1980
QY 1981 SLTSATSTSNIOSASQPRPQOGVYKLTMAOLTQLTQGHGNOGLTVVIOGOGTQOL 2040
Db 1981 SLTSATSTSNIOSASQPRPQOGVYKLTMAOLTQLTQGHGNOGLTVVIOGOGTQOL 2040
QY 2041 QLIPOGVTVLPFGGQOLMQAMPNGTVQRFLETPLATTTATTTATTTTAAAGTGEQR 2100
Db 2041 QLIPOGVTVLPFGGQOLMQAMPNGTVQRFLETPLATTTATTTATTTTAAAGTGEQR 2100
QY 2101 QSLTSFOMOVHODKTLPRPOSSSVGPAKAPOTAPPSARPOTOPSPAPPEVOTQPEV 2160
Db 2101 QSLTSFOMOVHODKTLPRPOSSSVGPAKAPOTAPPSARPOTOPSPAPPEVOTQPEV 2160
QY 2161 QYQTVSSHVPSSEAPQTHAQQSKPYAAQSQPSNVQGSQPVAVQSPSOTRIRPSTPSQL 2220
Db 2161 QYQTVSSHVPSSEAPQTHAQQSKPYAAQSQPSNVQGSQPVAVQSPSOTRIRPSTPSQL 2220
QY 2221 SPQOQSOVOTTTSQPIPIQPHSTLQIPSGQSPQSPQVOSSTQTLSSGQTLNOVSVSPS 2280
Db 2221 SPQOQSOVOTTTSQPIPIQPHSTLQIPSGQSPQSPQVOSSTQTLSSGQTLNOVSVSPS 2280
QY 2281 RPOLOIQOPPOPIYAPLOLQOQOVLSQIOSOVYAOIQAOQSGVPOQIKQLPIQIQSS 2340
Db 2281 RPOLOIQOPPOPIYAPLOLQOQOVLSQIOSOVYAOIQAOQSGVPOQIKQLPIQIQSS 2340
QY 2341 AVQTHQIQOVVTVQAASVQEQLOQVQOLRPOQOQKKQOQEIKEHETLQASNOSEIIOQK 2400
Db 2341 AVQTHQIQOVVTVQAASVQEQLOQVQOLRPOQOQKKQOQEIKEHETLQASNOSEIIOQK 2400
QY 2401 VVKKHNAVIEHLKQKSKMTPAEREENORMIVCNQVMKYIILDKIDKEEKOAAKKRKEESV 2460
Db 2401 VVKKHNAVIEHLKQKSKMTPAEREENORMIVCNQVMKYIILDKIDKEEKOAAKKRKEESV 2460
QY 2461 EOKRSQONTKLSALFHKKEOLRAELILKKRALDLKDLOIEVOEELKRDILKIKKEDLMQ 2520
Db 2461 EOKRSQONTKLSALFHKKEOLRAELILKKRALDLKDLOIEVOEELKRDILKIKKEDLMQ 2520
QY 2521 LAQATVAAPPCPVTPLVLPAPAPAPSPPPPGVQHTGLSTPLTPVASQKRKEEKDS 2580
Db 2521 LAQATVAAPPCPVTPLVLPAPAPAPSPPPPGVQHTGLSTPLTPVASQKRKEEKDS 2580
QY 2581 SSKSKKKKMIISTSKTKKDTKLYCICKTPYDESKFYIGCDRCQNNYHGRVGILOSEAE 2640
Db 2581 SSKSKKKKMIISTSKTKKDTKLYCICKTPYDESKFYIGCDRCQNNYHGRVGILOSEAE 2640
QY 2641 LIDEVYCPOCQOSTEDAMTVLPLTEKDYEGIKRVLSIAQHKAMPLEFVNDNDAPDY 2700
Db 2641 LIDEVYCPOCQOSTEDAMTVLPLTEKDYEGIKRVLSIAQHKAMPLEFVNDNDAPDY 2700
QY 2701 GVIKPEMDLATEMERVQRRYKTEFVADMTKIFDNCRYYNPSDSPEYQCAVLESFEV 2760
Db 2701 GVIKPEMDLATEMERVQRRYKTEFVADMTKIFDNCRYYNPSDSPEYQCAVLESFEV 2760
QY 2761 QKLGFKASRSHNNKLQSTAS 2781
Db 2761 QKLGFKASRSHNNKLQSTAS 2781

RESULT 2
AA57452
ID AA57452 standard; Protein: 2907 AA.
XX
XX AA57452;
XX
XX 22-FEB-2000 (first entry)
XX
XX Human transcriptional regulatory factor SEQ ID NO:1.
XX
XX Human: transcriptional regulatory factor; TCOA1; BLAST detection;
XX
XX bromo-domain; cell proliferation; cancer.
OS Homo sapiens.

XX
PN WO957143-A1.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99MO-JP02340.
XX
PR 30-APR-1998; 98JP-0137631.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
XX
XX MPI: 2000-052940/04.
DR N-PSDB: AA39032.
XX
PT Transcriptional regulatory factor containing a bromo domain and gene
XX TCOA1 encoding it -
XX
XX Claim 1; Page 48-68; 154pp; Japanese.
CC The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulation mechanism. It binds
CC to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
SQ Sequence 2907 AA:
Query Match 98.9%; Score 14170; DB 21; Length 2907;
Best local Similarity 95.1%; Pred. No. 0;
Matches 276; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
QY 1 MYSEEEEDGDAEFTQDSDEDEDEEDDDSDYPEEMEDDDDASTESSFNSHST 60
Db 1 MYSEEEEDGDAEFTQDSDEDEDEEDDDSDYPEEMEDDDDASTESSFNSHST 60
QY 61 YSTPGRRKRVHRPSPILKEKDIPLEPRKSSDMLVPMNEHIMVATIEVLRNGTV 120
Db 61 YSTPGRRKRVHRPSPILKEKDIPLEPRKSSDMLVPMNEHIMVATIEVLRNGTV 120
QY 121 LRSPREFEDFCAALVSOEQCTLAEMHVVLLKAVLREEDTSMTTGSPADLKDSVNSTLY 180
Db 121 LRSPREFEDFCAALVSOEQCTLAEMHVVLLKAVLREEDTSMTTGSPADLKDSVNSTLY 180
QY 181 FIDGMTPEVLRVYCESDKENHVLRYQEAEDYRYGVEVNIKIVLOFLVDQFLTTNARE 240
Db 181 FIDGMTPEVLRVYCESDKENHVLRYQEAEDYRYGVEVNIKIVLOFLVDQFLTTNARE 240
QY 241 ELMSSEGVIOYDDHCRVCHKIGDLLCCEYSAVYHLCVPRLEVEDEDEQCCVCAHNV 300
Db 241 ELMSSEGVIOYDDHCRVCHKIGDLLCCEYSAVYHLCVPRLEVEDEDEQCCVCAHNV 300
QY 301 PGVTDCAVEIQKKNPYIRHEPIGYDSRRKRYWFLNRLIIEEDTENENKKIYYSTKYQ 360
Db 301 PGVTDCAVEIQKKNPYIRHEPIGYDSRRKRYWFLNRLIIEEDTENENKKIYYSTKYQ 360
QY 361 LAELIDLDKDYWEAEELKLEEMREIHRHMDITEDLTNKGASNKSFLAANEELIS 420
Db 361 LAELIDLDKDYWEAEELKLEEMREIHRHMDITEDLTNKGASNKSFLAANEELIS 420
QY 421 IRAKKGDIDNVKSPETEEKDKNETENDSKDAENKREEFQSLSEKSDDKTPDDPEQK 480
Db 421 IRAKKGDIDNVKSPETEEKDKNETENDSKDAENKREEFQSLSEKSDDKTPDDPEQK 480
QY 481 SE----- 482
Db 481 SEPTVEGDKGNSVSNLGDNTNATSEETSPSEGSVPVGCISPTDSSNMAEKYASEL 540
QY 483 ----- 482

Db 541 PDVPEEPNKTCSSNNTSATTTSIQPLSENSSSSELNSSQSSAKAADDPENGERSHT 600
QY 483 -----VDFKSEKNGELSPGAGKASGSTRITTLRNPDSKLSOLKSOQVAAA 534
Db 601 PVSIOEBIVDFPSEKSTGELSESPGAGKASGSTRITTLRNPDSKLSOLKSOQVAAA 660
QY 535 HEANKLEKEKEEVLVNSQOEISRLSTPKREVMKGNINNFKLQOEKRYVYHNOYSTNS 594
Db 661 HEANKLEKEKEEVLVNSQOEISRLSTPKREVMKGNINNFKLQOEKRYVYHNOYSTNS 720
QY 595 PALNKHORHDHDKRRHLAKHFCITPAGEFRKMGSVHGSVLTSTLTTLTTOLENNIPS 654
Db 721 PALNKHORHDHDKRRHLAKHFCITPAGEFRKMGSVHGSVLTSTLTTLTTOLENNIPS 780
QY 655 SPLHPNASHRAMWIKAVQCSKPREFALALATLECAVKNVVMPLTIRELGHTRHMR 714
Db 781 SPLHPNASHRAMWIKAVQCSKPREFALALATLECAVKNVVMPLTIRELGHTRHMR 840
QY 715 SIEREEKYKKEKKEKOEETMOQATWVKTPEVKHQVWKOKGEERYVATGYGGMWSIK 774
Db 841 SIEREEKYKKEKKEKOEETMOQATWVKTPEVKHQVWKOKGEERYVATGYGGMWSIK 900
QY 775 THYRFPKPLPGNTNVYRKSLSEGTKNNDNNDNDESKRCSNPKIKITIEPSEKDEV 834
Db 901 THYRFPKPLPGNTNVYRKSLSEGTKNNDNNDNDESKRCSNPKIKITIEPSEKDEV 960
QY 835 GSDAKADONENDISKITEKDDOVKELDSDSKPCKEEMVEDDMKTESHVNCQES 894
Db 961 GSDAKADONENDISKITEKDDOVKELDSDSKPCKEEMVEDDMKTESHVNCQES 1020
QY 895 SOVDVNVVSEGFHLRTSYKKTSSKLDGLLERRIKOFTLEEKORLEKIKLEGKIGIK 954
Db 1021 SOVDVNVVSEGFHLRTSYKKTSSKLDGLLERRIKOFTLEEKORLEKIKLEGKIGIK 1080
QY 955 TSTNSKNLSRSPYTAKGCGCOSMSKROSPNANNDOPEDLIOGQSOSVSLRMSDP 1014
Db 1081 TSTNSKNLSRSPYTAKGCGCOSMSKROSPNANNDOPEDLIOGQSOSVSLRMSDP 1140
QY 1015 SHTNKLTPKDRVLDDVIRSPEKCPKONSIENTDIEEKVSDLASRQOEPTKSTKNDP 1074
Db 1141 SHTNKLTPKDRVLDDVIRSPEKCPKONSIENTDIEEKVSDLASRQOEPTKSTKNDP 1200
QY 1075 FIDDSKLASADDTGTLCKNNKPLIOEESDTYSSSKSALHSSVPKSTNDRDATPLSRAM 1134
Db 1201 FIDDSKLASADDTGTLCKNNKPLIOEESDTYSSSKSALHSSVPKSTNDRDATPLSRAM 1260
QY 1135 DFEKLCDCDSNSTLENSSDTVSIODSSEDMIVONSNESISQOFTREDOVAVLEPLK 1194
Db 1261 DFEKLCDCDSNSTLENSSDTVSIODSSEDMIVONSNESISQOFTREDOVAVLEPLK 1320
QY 1195 CEIVSGESTGNCEDRLPVKGTETANGKRPQOKLEERPVKCSDQIKLKTNTDKNNENR 1254
Db 1321 CEIVSGESTGNCEDRLPVKGTETANGKRPQOKLEERPVKCSDQIKLKTNTDKNNENR 1380
QY 1255 ESEKKGORTSTFOINGKDNKPKITLKGECLEKESRSRVSGNVKPKNNINIKIIPENDIK 1314
Db 1381 ESEKKGORTSTFOINGKDNKPKITLKGECLEKESRSRVSGNVKPKNNINIKIIPENDIK 1440
QY 1315 SLUTKESAIRPFIINGDVIEMEFNERNSEFSSHLLSSDAEGNRDSELTLPSTKESDST 1374
Db 1441 SLUTKESAIRPFIINGDVIEMEFNERNSEFSSHLLSSDAEGNRDSELTLPSTKESDST 1500
QY 1375 QTTTPSASCPEBSNVNOVEDMEIETSEVKKYVTSPLTSEBSNLSDNFIDENGLPIKNE 1434
Db 1501 QTTTPSASCPEBSNVNOVEDMEIETSEVKKYVTSPLTSEBSNLSDNFIDENGLPIKNE 1560
QY 1435 NVNSESRRKYITTEVTMTSTVATESTVIKVEKGDQOTVYSSSTENAKSVTTTTTIVT 1494
Db 1561 NVNSESRRKYITTEVTMTSTVATESTVIKVEKGDQOTVYSSSTENAKSVTTTTTIVT 1620
QY 1495 KLSPSTGGSVDLISVKEOSKTVTTVTBDSLTGGLVTSMTVSKKEYSTRDQVVKLMKF 1554
Db 1621 KLSPSTGGSVDLISVKEOSKTVTTVTBDSLTGGLVTSMTVSKKEYSTRDQVVKLMKF 1680

QY 1555 SRBKTRSGTALPSYKREFTVKSTKSTIFVLPNDLKLAKKGGIREVPYNNYNAKPALDI 1614
Db 1681 SRBKTRSGTALPSYKREFTVKSTKSTIFVLPNDLKLAKKGGIREVPYNNYNAKPALDI 1740
QY 1615 WPYSPRPFGITWRYRLQTVKSLAGVSLMLRLMLASLRWDDMAAKVPPGGGSTRTETSE 1674
Db 1741 WPYSPRPFGITWRYRLQTVKSLAGVSLMLRLMLASLRWDDMAAKVPPGGGSTRTETSE 1800
QY 1675 TEITTTTIIKRVGVPRGIFEFYCIRKILICPIGVPEPKTPTTPQKGLRSSALRKRP 1734
Db 1801 TEITTTTIIKRVGVPRGIFEFYCIRKILICPIGVPEPKTPTTPQKGLRSSALRKRP 1860
QY 1735 TPQOTGVIIETVWAEELMEIEMELRAFAEVEKEKAQAEQAKKRLDOOKPVIATSTT 1794
Db 1861 TPQOTGVIIETVWAEELMEIEMELRAFAEVEKEKAQAEQAKKRLDOOKPVIATSTT 1920
QY 1795 SPTSSTSTISPAQKVVAVDSSGVTTGTIKMVLTKVGSFATVTFQONKNEHOTFAWYK 1854
Db 1921 SPTSSTSTISPAQKVVAVDSSGVTTGTIKMVLTKVGSFATVTFQONKNEHOTFAWYK 1980
QY 1855 QGOSNSGVVOOQKVLGILPSSGTSGOFTSPOPRTATVTRPNTSGSGGTTNSGOVIT 1914
Db 1981 QGOSNSGVVOOQKVLGILPSSGTSGOFTSPOPRTATVTRPNTSGSGGTTNSGOVIT 2040
QY 1915 GPQIRPGMTVIRTPLOOSTLGKAIIRTPVWVQGPAPQOVFTQIIRQOPVSTAVSAPNTVS 1974
Db 2041 GPQIRPGMTVIRTPLOOSTLGKAIIRTPVWVQGPAPQOVFTQIIRQOPVSTAVSAPNTVS 2100
QY 1975 STPGOKSLTATSTSNISQSSASQPPRPQOQOVKLTMAQLTQLVQHGGNGGLTVVIOGQ 2034
Db 2101 STPGOKSLTATSTSNISQSSASQPPRPQOQOVKLTMAQLTQLVQHGGNGGLTVVIOGQ 2160
QY 2035 QTTGOLILIPQGVTVLEPGPOQLMAQAMPNGTQVORELFTPLATTAATSTTTVSTAA 2094
Db 2161 QTTGOLILIPQGVTVLEPGPOQLMAQAMPNGTQVORELFTPLATTAATSTTTVSTAA 2220
QY 2095 GTGEBORSKLSPOWQVHQDKTLPPAOSSSVGPAKAOPQTAOPSARPOQTOPOSPAPQEV 2154
Db 2221 GTGEBORSKLSPOWQVHQDKTLPPAOSSSVGPAKAOPQTAOPSARPOQTOPOSPAPQEV 2280
QY 2155 QTOPEVOTOTTVSSHVSEKOPTHAOSKPOVAAQOSPOGNSVGOQSPVRVQSSQTRIRP 2214
Db 2281 QTOPEVOTOTTVSSHVSEKOPTHAOSKPOVAAQOSPOGNSVGOQSPVRVQSSQTRIRP 2340
QY 2215 STPSQSLSPGOQSOVOTTSOPIRPIQPTSLQIPSOQPOSOPOVOSSTQTLSSGOTLNOV 2274
Db 2341 STPSQSLSPGOQSOVOTTSOPIRPIQPTSLQIPSOQPOSOPOVOSSTQTLSSGOTLNOV 2400
QY 2275 SVSSEPRPOLIQPOPOVTAIVPOLQOQOVYLSQIOSQVVAQIOAQOQSGVPPQITKLOLP 2334
Db 2401 SVSSEPRPOLIQPOPOVTAIVPOLQOQOVYLSQIOSQVVAQIOAQOQSGVPPQITKLOLP 2460
QY 2335 QIOQSSAVOTHOJONVTVVAAASVOEOLQVVOOLRDQOQKKKQOQIEIKREHFLQASNOS 2394
Db 2461 QIOQSSAVOTHOJONVTVVAAASVOEOLQVVOOLRDQOQKKKQOQIEIKREHFLQASNOS 2520
QY 2395 EIIOQOVYMKHNAVIEHLKOKKSWTPAREBENOMIVCNOVMYILDKIDKEEQAAKRR 2454
Db 2521 EIIOQOVYMKHNAVIEHLKOKKSWTPAREBENOMIVCNOVMYILDKIDKEEQAAKRR 2580
QY 2455 KREESVEQKRSKONATKLSALLFKHKQOLRAELIKKRALDLKLOLIEVOBELKRDLIK 2514
Db 2581 KREESVEQKRSKONATKLSALLFKHKQOLRAELIKKRALDLKLOLIEVOBELKRDLIK 2640
QY 2515 EKIDMOLAQATAVAAPCPVTPVLPAPAPPPSPPPGVOHGLSTPLTPVYASQKRR 2574
Db 2641 EKIDMOLAQATAVAAPCPVTPVLPAPAPPPSPPPGVOHGLSTPLTPVYASQKRR 2700
QY 2575 EEEKDSSSKKKKKMISTSKETKKDKTKLYCICKTPYDESKFYIGDCRCOMWYHGRGVGI 2634
Db 2701 EEEKDSSSKKKKKMISTSKETKKDKTKLYCICKTPYDESKFYIGDCRCOMWYHGRGVGI 2760

QY 2635 LQSEALIDDEVYCPQOSTEDAMTVLTPTLEKDEGLKRVLSLQAHKAMPLEPEVDN 2694
|||||
Db 2761 LQSEALIDDEVYCPQOSTEDAMTVLTPTLEKDEGLKRVLSLQAHKAMPLEPEVDN 2820
QY 2695 DAPDYGVIKPEPMDLTMERVORRYEKLTEFVADMTKIFDRCRYNSDSPEYOCAY 2754
|||||
Db 2821 DAPDYGVIKPEPMDLTMERVORRYEKLTEFVADMTKIFDRCRYNSDSPEYOCAY 2880
QY 2755 LSEFVQKLKGFKASRSHNNKLOSTAS 2781
|||||
Db 2881 LSEFVQKLKGFKASRSHNNKLOSTAS 2907

RESULT 3
AAB94078
ID AAB94078 standard; Protein; 557 AA.
XX
AC AAB94078;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14273.
XX
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14273; 2537bp + CD ROK; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides, and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH1742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX
SQ Sequence 557 AA;
Query Match 18.8%; Score 2695.5; DB 22; Length 557;
Best local Similarity 79.0%; Pred. No. 1; 6e-130;
Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;

QY 1811 MVAPIGSVTTGKRVLTTRKVSAPATVTFQONKFNQTFATVWKGOSNGVVOQKVL 1870
|||||
Db 1 MVAPIGSVTTGKRVLTTRKVSAPATVTFQONKSFQTFATVWKGOSNGVVOQKVL 60
QY 1871 GLIPSSGTSGQTFSTFQPTATVTRPNTSGSGTSSNSQVITGPQIRPGATVIRTPDQ 1930
|||||
Db 61 GLIPSSGTSGQTFSTFQPTATVTRPNTSGSGTSSNSQVITGPQIRPGATVIRTPDQ 120
QY 1931 QSTLGRKAIIRTPVAVPGAPQVMTQIIRGPVSTVAVSAPNTVSTPQKSLTSATSTSN 1990
|||||
Db 121 QSTLGRKAIIRTPVAVPGAPQVMTQIIRGPVSTVAVSAPNTVSTPQKSLTSATSTSN 180
QY 1991 IOSSASQPRPQOGQVYKLTMAQLTQLTQGHGNGQLTVVIOGOGQTTGQLIPQGVTVL 2050
|||||
Db 181 IOSSASQPRPQOGQVYKLTMAQLTQLTQGHGNGQLTVVIOGOGQTTGQLIPQGVTVL 240
QY 2051 PGPQQLMQAAMPNGTVORELFTPLATTAATTAATTTTAACTGAGTGEOROSKLSFQMGV 2110
|||||
Db 241 PGPQQLMQAAMPNGTVORELFTPLATTAATTAATTTTAACTGAGTGEOROSKLSFQMGV 300
QY 2111 HQDKTLPPAOSSSVGPAPAKAPQTAAPSARPPQTPQSPAPQPEVQTPVQVOTVSSHV 2170
|||||
Db 301 HQDKTLPPAOSSSVGPAPAKAPQTAAPSARPPQTPQSPAPQPEVQTPVQVOTVSSHV 360
QY 2171 PSEAOPTHAOSSKPOVAAOSQPOSNNVQOSPVRVQSPQTRIRPSPQSLSPQOSQOVQT 2230
|||||
Db 361 PSEAOPTHAOSSKPOVAAOSQPOSNNVQOSPVRVQSPQTRIRPSPQSLSPQOSQOVQT 420
QY 2231 TTSQPIPIQPHSTLQIPSQGQPOSQPOVQSTQTLSSGQTLNQVSVSPSPQLQIQPQ 2290
|||||
Db 421 TTSQPIPIQPHSTLQIPSQGQPOSQPOVQSTQTLSSGQTLNQVSVSPSPQLQIQPQ 446
QY 2291 POVIAVPQLQOQVQVLSQIOSQVVAQIOAQOQSGVPOQIKLPIQIQOSSAVQTHQIQNV 2350
|||||
Db 447 ----- 446
QY 2351 VTVQAASVQEQQLRVQQLLQDQOQKKQOQEQIKREHTLQASNSQELIQQVYKHNATV 2410
|||||
Db 447 -----QVYKHNATV 457
QY 2411 HLKOKKSMTPAEREENQRMIVCNQVKKYILDKIDKEEQQAQKKRREESVEQQRKQNT 2470
|||||
Db 458 HLKOKKSMTPAEREENQRMIVCNQVKKYILDKIDKEEQQAQKKRREESVEQQRKQNT 517
QY 2471 KISALLFKHKEQLRAEILKKRALLDKDLQIEVQELKRD 2510
|||||
Db 518 KISALLFKHKEQLRAEILKKRALLDKDLQIEVQELKRD 557

RESULT 4
AAB68735
ID AAB68735 standard; Protein; 976 AA.
XX
AC AAB68735;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 32997.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL12838.
 PT
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
 XX
 PS Disclosure; SEQ ID NO 32997; 21bp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 976 AA;
 Query Match 10.9%; Score 1557.5; DB 22; Length 976;
 Best Local Similarity 36.4%; Pred. No. 1.3e-71;
 Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;
 QY 4 EEEEDGDAAETODSDEDEDEDEDDDDSDYPEEMEDDDDDSYCTESSFRSHSTYSS 63
 DB 99 ESEHYGSDGSDSESDSDNEDMDLTPSDDESLEVANESSEFSVC---SEFONGV--- 152
 QY 64 TPGRRKPRVHRPRSPILKE-KDIPLEFPKSSEDLWPNNEHINNVIAIYEVNFGTVLR 122
 DB 153 ---GRPPRPSPRPVWLOEGROYAALDIPDSEDLFIANTHVLAALSTYELRFRHMVR 209
 QY 123 LSPRFEDFCALVYSQEOCTMAEMHVLKAVLREBDTSNTTGGPADLKDVSNTLYFT 182
 DB 210 LSPRFEDLCALACEQSALTVEVHMLKALIREBDAQCTHFGPDKDXTVNISLYLI 269
 QY 183 DGMWPEVLRYYCESDKEX---HHVLRYOEAEDYRGVGNKIKVLOFVDOFLTNIA 238
 DB 270 DSTWPEVLRYSVDSKTFEDRNPFILSHTE---YPTGIDNRLEVLOFSLDSOFLTSISI 326
 QY 239 REELMSEGVIOYDHCRCVCHKLDLCCETSAUYHLECVKPLPEVPEDEMOECVAVH 298
 DB 327 RDVHLOEGRIHYDHCRCVCHKLDLCCETCPAYHLECVPRPNNDVPTEDMOGGLCRSH 386
 QY 289 KVPGVTCVAETQKKRPYTRHEPTGYDRSRKKYFNLNRLIIEEDTENENKKTWYSTK 358
 DB 387 KVGSGVDCVLFOERKOGVLIHRDSSLGVDRHGRKKYFIARIRFI-BDOENFT---CMYYSTT 442
 QY 359 VOJAEIDLCIDKDYWEALCKLIEEMREIHRHMDITEDLNNKRGSKSLAANETIL 418
 DB 443 SKLKLLSRDLAELETRLSQITERKDELRQMKLTETLTNEKHKRSYI----- 494
 QY 419 ESIRAKKGDIDNVASPEETEKDKNETENDSKDAEKNREFEQDSLEKSDDKTPDDPEQ 478
 DB 495 -----EIEGEAKN-----ELLEKEVLDDEDEKO----- 516
 QY 479 GKSEYGDGFKSEKNGELSESFGACKGASGSTRITITRLRNPDSKLSQKSOQVAAAHAN 538

DB 517 -----GDAKSESQISIE-----GTCKOECKMWTIR-----QKSNQLT----- 547
 QY 539 KLEFKEBEVLVYNVSGOEISRLSTKKREVIKMGNNINNEFKLGOGEKRYVYNOSTNSFALN 598
 DB 548 -----NGLTFEKLGMEOGFEKNYNOYSTNFIALN 576
 QY 599 KQHREDHDKRRHLAHKFCPLTPAGEFKWNGSVGSKVLTISTRLTLITOLENNIPSSFLH 658
 DB 577 KPORNEERDKRRLSHKFSILTASDFKWIJMTGTDNNITTLRQTLINFEENIAASFILN 636
 QY 659 PNNASHRANWKAVQMCSPREPALALILECAVKKYVWMLPIWREPLGTRHLHMTSIR 718
 DB 637 INNVVNRKKIWNNAVWNAARRSEFAVVLTLFOASLKSVPANWHEOLGHTTLQRTISAER 696
 QY 719 EEEKKKKKKKKKKO---EEETMQOATWVKTFFPVKHQVWKOKGEEYRVYGGMSWISKT 775
 DB 697 EERKKLEKREKRRRDEEENRLAFNYIKTYTLGLKIQVWKOKGEEYRVYGGMSWMLSS 756
 QY 776 -----HYREFVPKLPNTNVNVRKSLSEGTKNNDENWDESDKRKCSRPKKIKIEPDSER 830
 DB 757 RRCGVARRAQAQPLTHNRVYVHYTM---GEENDVNEIT-----LVDRPTORFMOQCES 805
 QY 831 DEYKGSDAKGAQO---NENDISKITEKKODVKEILDSDSKRCKREPEHVEDDMKTESH 888
 DB 806 SNVDGVCYHLPDQYKNVYIEDVTE-----KIKKH 836
 QY 889 VNQOESSQVNVVNSGCFHL--RTSYKKKTSSKLDGLERRIKQPTLEKOKLEKIKLE 946
 DB 837 -----IDYSKALNAPGRTIYSKVARKRSDLDLDRUKLAIVEEOM---ASKIP 882
 QY 947 GGIGIGIGTSTNSSKNISES 966
 DB 883 SDMKPLIVSSQNTWANSKQT 902
 RESULT 5
 AAB54364
 ID AAB54364 standard; Protein; 328 AA.
 XX
 AC AAB54364;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
 XX
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW noctropic; immunomodulatory; relaxant; contraceptive; gynecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05989.
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-579444/54.
 DR N-PSDB; AAC99129.
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing, treating, or ameliorating a medical condition, particularly pancreatic

XX BLANCHARD AD, Champion B, Page KR;
PI WPI: 2002-306794/35.
DR N-PSDB; ABR47549.
XX
PT New SKAT-2 polypeptide transcription factor, useful for treating or
PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
XX acid, antibodies and modulators
XX
PS Disclosure: Page 37-38; 47pp; English.
XX
CC The invention describes an isolated SCAM/KRAB protein associated with a
CC Th2 phenotypic (SKAT-2) polypeptide (I). (I), especially where expressed
CC in recombinant cells, is used to identify agents that modulate SKAT-2
CC activity, and may be useful therapeutically. The agents, also (I) and
CC the polynucleotides (II) that encode it, are useful for treatment and
CC prevention of diseases responsive to SKAT-2 modulation, specifically
CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
CC Alzheimer's disease (and other neurological disorders, particularly where
CC related to apoptosis); conditions involving T-helper responses, and
CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
CC or humoral responses, to reduce development of a Th1 phenotype, and to
CC manipulate the immune response in autoimmune diseases, infections and
CC cancer. (II) is also used for recombinant production of (I); as source
CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
CC detection of mutations and for monitoring SKAT-2 expression in
CC association with disease. This is the amino acid sequence of the peptide
CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
CC SKAT-2.
XX
SQ Sequence 149 AA:
XX
Query Match 5.5%; Score 789; DB 23; Length 149;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2633 GIIQSAELIDYVPCOSTEDAMTVLPLETKDYEGKRLVRLSLQAHKMAFPLEPVD 2692
DB 1 GIIQSAELIDYVPCOSTEDAMTVLPLETKDYEGKRLVRLSLQAHKMAFPLEPVD 60
QY 2693 PNDAPYGVYIKPEMDLATEERVORRYEKLTEFVADMKIKIDNCRYYNPSDSPFYQCA 2752
DB 61 PNDAPYGVYIKPEMDLATEERVORRYEKLTEFVADMKIKIDNCRYYNPSDSPFYQCA 120
QY 2753 EYLESEFVQKLGFKASRSNNKLOSTAS 2781
DB 121 EYLESEFVQKLGFKASRSNNKLOSTAS 149
XX
RESULT 9
AAU31944
ID AAU31944 standard; Protein: 246 AA.
XX
AC AAU31944;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2435.
XX
KW Human: vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX

PR 26-JAN-2001; 2001US-0770160.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
DR WPI: 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
XX
PS Claim 20; Page 529-530; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or
XX physiological interactions of the polypeptide. Vectors comprising
XX the nucleic acids encoding the polypeptides and cells genetically
XX engineered to express them are also useful for producing the proteins.
XX The proteins are useful in genetic vaccination, testing and
XX therapy, and can be used as nutritional supplements. They may be used to
XX increase stem cell proliferation; to regulate haematopoiesis; and in
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX immune suppression and/or stimulation; as anti-inflammatory agents; and
XX in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 246 AA:
XX
Query Match 5.4%; Score 781; DB 22; Length 246;
Best Local Similarity 67.8%; Pred. No. 1.6e-32;
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;
QY 751 HQVWKQGEERYVYGVGWSMISKTHYRFVPLPGNTNRYNRSLEGTNNNDENDES 810
DB 7 HQVWKQGEERYVYGVGWSMISKTHYRFVPLPGNTNRYNRSLEGTNNNDENDES 66
QY 811 DKRCKSPKIKIIEPSEDEYKGSAAAGADQNEIDIKTEKKDDYKELLDSDSK 870
DB 67 DKRCKSPKIKIIEPSEDEYKGSAAAGADQNEIDIKTEKKDDYKELLDSDSK 126
QY 871 PCKEPEVDDDKTESHVNCQSSQVD--VVNVSEGFHLRTSYKKKTSKLDGLDER 928
DB 127 PCKERTN--GSRXHXERVYCKLSGEFSSRCQCXXCFSSKDXLQKENKITQIRWTSXKE 184
QY 929 IKQFTLEKQRLKIKLEGIGIKGTSTNSSRNLSSEPVITRAKECCQSDSRQROSP 987
DB 185 NXYVYVYGRKATATRRKNQYGMNXXGRKQSTNSSKNLSSEPVITRAKECCQSDSETRPE 243
XX
RESULT 10
ABB62887
ID ABB62887 standard; Protein: 997 AA.
XX
AC ABB62887;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SPQ ID NO 15453.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX

PT interactions -
XX Dislosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent.
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB85737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2768 AA;
SQ
Query Match 4.0%; Score 572.5; DB 22; Length 2768;
Best Local Similarity 18.4%; Pred. No. 2.7e-20;
Matches 562; Conservative 437; Mismatches 1147; Indels 905; Gaps 128;
QY 78 PLEEKDIP---PLEPKSEDLMPNEMIMVIAIYELRNFGYLRSPREFDFCA 134
DB 84 PITTCEPVVDVPHYSPEGTSLIPEK---GCSIEKFPBGACVPSNPKPCELCYC 140
QY 135 LVSOECSTLMAEMHVLLKAVLREEDTSNTTEGPAADKDSVNSTLYFDGMTPEVLARY 194
DB 141 INNGTKVQWQ-----BCTLH-VQG-----CLPTIY 163
QY 195 GCSDEYHNHLYQEAEDYPTGPNVKKIVLOFLVDQILTN----- 236
DB 164 NNGS-----CCCVRYSCDH-----ENELD-----FMDSTTTTIVRPTTGFLASTMP 209
QY 237 -----IAEELMSGV-IQYDHCYVCHL-GDLIC-----TCSAVYHL 275
DB 210 PTTTDCIHGELFADASLCKGNACCHCKCMRGDIYCAVQCEVPMMAANGSCRMPPA 269
QY 276 ECVKRPLEVPEDMOCEVCAHKGVTDCV-----AEIOKNKPIIRHEPIGYDRSR 329
DB 270 BECCPSNVVCDSDSTTEIVETTESATSVPAKGIHAEIKEDVDLDDDKNKE 329
QY 330 KTWFLNRRIITEDTENENEKIKIWIYSTVQLAELIDCDKDTW---EALCKILEMRE 386
DB 330 TATISAEIGSGEIEEIEEEKD---KATVAVPOVTD-EKDFSEFESSTAGIPDSR- 383
QY 387 ELHRHMDITEDLTKNARGSKSFLAANEILESTRAK-----GDDDNK----- 432
DB 384 -----IDLPSSTEEKSSTEAEDIVKIVTTPPEGSGEEDVPRPSQIPEKIT 434
QY 433 -----SPEE-----TEKDKNET-----ENDSKD 450
DB 435 EDELKIVSTASAKASPEEEVYATATTSAPTEEDVAKPTAGTISEEEBEKPTPAEGSGE 494
QY 451 AEKN-----REFEDQ-----SLEKSD-----DKTPDDPEQKSEV 483
DB 495 EEKDVKVTAPAEETDEAKPTSAVASEKEQEPKPSGSGDELDLKPPTAPATGATSA 554
QY 484 GDFKSEKSNGLSESP-----CAGKAGSSTIILRLRNPDSKLQKSOQA 531
DB 555 SESEEDQDGKSTEAPTSVDITIEAPKPTESSEASGEEDVAKETTPPAGEASIAEGEEIV 614
QY 533 AAHAHANKLFKGEKLV-----VNSQGEISRLST-----KREVIK 568
DB 615 KQTPPAGESSSGDEIEIVGTTPAEESSESEDELDLKVTPPAGEPSVAGEEELAKTTTA 674
QY 569 GINNNYFKLQEGKRYVYHNOYSTNSPALNKQHREDHDKRRHLAKFCLTPAGEKMWG 628
DB 675 GPPS---IAGEEIEIVK-----TTPAGESSIAEGEEIVK-----VTTPAGESSSEG 717
QY 629 SVHSGKVL-----ISLRLIITOLENNIPSPFLHPNASHR 665
DB 718 EBEITKVTTPAGESSESGDEIEIVKSTPAGEPIISGEEDVIAKATYAPKSDIEGVKEPET 777

QY 666 ANWIKR--VQMSCKPREFALALALECAVKNPVYMLPIRMREFLGHTRLHRTSIEREK 723
DB 778 ATEVPAEEVEDPAK-----TTPIAEEEPPIAGTPIPDGIS-----GEEIEIVK 822
QY 724 VKKEKKQEEETMQOATVKTTPPYKHQWKKQGEERYTGTGCMWSIKTHVYFVK 783
DB 823 GTTPQLEQOPELISESTEV---PVAED-----D 848
QY 784 LGNTNVN-YRKSLESTKNMDEENDESDKRCOSRPPKIKIIEPDEKD----- 831
DB 849 LSSSTISASAIASSTBQVQASSETTSSAPARAGDKDEAATVTPPADKDEEYEDATDL 908
QY 832 ---EVKSDAAGADONEMDISKITEKDDODYKELDSDPKCKEPEVDDMK----- 884
DB 909 PVEDVQSTTAKTTTTEQKKEESTTEADAEIEVTTSSPADQVEAPADKDHDEED 968
QY 885 -----TESHVNCOESSQVDVNVVSGFHL-RYSYKKTYSKFLDGL 924
DB 969 VQATADLPKSDIGPPVVDTEATTGQPTSDETATDKPPSVYLLPVYQEVPSSTAKVDNR 1028
QY 925 LEERRIQFTL-----EKKORLEKILEGGIG-----IGKTSNKSNLSPVITKA 972
DB 1029 NDETEKPTLPSPGSDQSSPLPAMDLPAGIREGDCLVGKTYANNVTVPATAPCDVSC 1088
QY 973 KEGCOSDSMKROESPANNDQPE-----DLIOGCSQS---DSSVLRMSDPSTHTKLYP 1023
DB 1089 K-CISLIVACQOMEKLPENLEKCTVAADLDGCCPTIYCEBSTSAKDEESTAK--P 1144
QY 1024 KDRVLDVSTRPEKCPKONSIENTDIERK-----VSLASRGQEPYKSKTK 1070
DB 1145 DNKIDVSEISTE-BIPDVIIMPITGITEQPLSHVKNPDEIOPVTSVPAQFESTAKVD 1203
QY 1071 GDNFIDDSKLASADIGLICKNKKPLIOESDITVSSSKSALHSVSKSTNDRAPPL 1130
DB 1204 KRP--IDES-----AEKPKPIGESED-----SKPIDESEDKKVV 1237
QY 1131 SRAMDFEKLGDSESNSTLENSDPTVSIOSSSEDMIV-----QNSNESISEQFTR 1183
DB 1238 E-----ESAEDKKPVDESEKPKPLPTVPAISEIKESKPEDEKKT 1278
QY 1184 EOD---VEYLEPKCELYSGESTGNCEDRLPYKGTGANK---KPSQOKLEERPYNKC 1236
DB 1279 EADFAAPTEQPEATPPAQIADTAEKEDVDLKTATSAVPSGDELBKADKKRTE----- 1332
QY 1237 SQOI---KLKNTDCKNNENR-----ESEKKQRTSP--FOJNKQKPKIYLKGECLKEI 1287
DB 1333 TAOIPDAETPASTDEPESSTELPTVLDKKPEEDSTKGTAPBSOKVPEVPTSASTENEI 1392
QY 1288 SESRVVSGNVEPKVNNINKIIP--ENDIKSLVKE-----SAIRPTINGDYI----- 1332
DB 1393 EESDKFTTYAPRKISASDELEPTAEEDLVPAFTEPILESFEVSTKRPVAGPPLPLARA 1452
QY 1333 -MEDFERNSETKSHLLSSDAEGNYRDSLETLPSTKESDTOTTP--SASCPESNVN 1390
DB 1453 OPEKKKPVDAETSTEADISTEPSAEVEKEASGETSESDNEDIDAGASSTPVPAVASEDKTPS 1512
QY 1391 QVEDMIEIETSVKKYVSSIT--SEESN---LSNDIDENGLPJ-----KNNEVNGE 1439
DB 1513 TEKTYE---ADDKFTVAALADDEESNLPKLPQDIFEEBA-PVAATTAAPSKDD--GE 1565
QY 1440 SK-----RKTVTEVTMTSVATATESKYIKKEGDKQTVVSTENCAKSYT 1487
DB 1566 OKPVEVEEKPRIEDGQKPIRDERSTPIS-----SENIEPESDRATIIASKEE--PSERS 1618
QY 1488 TTTTIVTKLSTSTSGSVDIISVKEOSKTVVTTVDSLTGGLVTSMTVSKYSTRD 1547
DB 1619 TGAPTKDEPAEESTIDAPESDES--KEPSESEVPTVAFA---GEKIPSTIIPDEEPAT 1673
QY 1548 KVKIMKFSRPPKTRSGTALPSYRKFTVTKSTKSIFLPLPDDLKLARKGCIREVFPFN 1607
DB 1674 SAPVAKPDEDVEKETSTELPTDAPASSEDEMS---STDQIPS-----EVP---E 1717

```

QY 1608 AKPALDIMPSPRPTEFGITWRYRLQTVKSLAGVSLMLRLMLASLRMDMAKAVPPG-- 1665
Db 1718 KRP-----LTPAOTPEBGD 1732
QY 1666 -GSTRETSETETTTTEIIRRDVPGYINFEYCKIRKIICPGVPTPKETPPQRK-GL 1723
Db 1733 VGATAAPTSTDEVPPVQRLPE-----EVLAIPIQPSTETGT 1768
QY 1724 RSSALRKRETPKQNGPVIIETWVAEEELMEIRAFARVERK--EKAQAVEQ-----Q 1776
Db 1769 KODETTAAPSIDRK-----EPYVEIDEAATVAPISRKDKPTREEKPVQOKPTGEE 1822
QY 1777 AKRLEQOKPTVIATSTSTSTSTSPISPAOKWMAPIGSVTTGKMLTTKVGSPAT 1836
Db 1823 PSEEEKEKPIEDOVSTEGVSTASAGSTES-----SEFKPSTEGEVAEKPED-- 1873
QY 1837 VTFQONKNFHQTATWVKGQOSNS-GVYVQVQKVLGIIPSTG--TSQOTFTSFQPRIA 1892
Db 1874 -----KQPSSTAQAAPVETIPEISTELPAQDDDKPTSEAPVDSDEDTSA 1916
QY 1893 TVTIR-PNTSGSGGTTNSQVITGPOAIRPGMTVIRTPLOOSTLGKAIIRTPVWVGAPQ 1951
Db 1917 PSDEKIPSVSG-----EVEGPEV-----TTASPOAAEBDELKTPASEPSTSD 1960
QY 1952 QV-MTQIIRGQPVSTAVSAPNTVSSTPGOKSLTSAVST-----SNTQSSASQPPRPQOQ 2005
Db 1961 KVPETEVOKEDETKADETESVATQVSDVATSTISAPVAGSDIEDEQATASP--EEEEE 2018
QY 2006 VKLTMAQLTQITQGHGNGQILTVVIGOGGOTGQQLIPQGVTVLP--GGGQQLMQ-AAM 2062
Db 2019 IKPTIAAPAELTPQ-----PSEKEPVDE-QEVESGTAKTAPAESDQPIDELIAPA 2065
QY 2063 PNGTVQRFLEPLTATATTAATTTTSTVSTAAGTGEOGROSKLSPOMVHODK-----TL 2116
Db 2066 TSGIDE-----ASTAPRKESTIVASAA-----SP-AVHDEIKDVTITQ 2106
QY 2117 PPAOSSSVGPAKAQPOTA-----QPSAR-PQPOQPOSPAQPEVQTOPEVQV--QTVV 2166
Db 2107 PVADEKEVAAPQODETKTSDIVSTSPTAODEKQDKTEAPATVATVSSPADSADSTP 2166
QY 2167 SSHAPE-----EAQPTH--AOSKRPQVA-AOSQPSNVGQSPVRY--QSSSORIRPS- 2215
Db 2167 TVEVPSPEVIDTKPMDIMSQTIAPHTADGAASTSTEDQAPVATVSPQDAEKTPVPAP 2226
QY 2216 -----TFSQLSPPGOOSQOVOTTSSQ-----PIPIQPHSLQIPISQOGQPOSPQOSSSTQ 2264
Db 2227 QDSQKTSSEAPQADDELPAATPLDNKIPATVAPQOTDQGVPAATAPLEDKIQTTAAR 2286
QY 2265 L-----SSGQTLNQVSVSSRPQLOIQOPQOVIAVPOLOQOVQVLSQIQSOVVAQIQ 2318
Db 2287 LDEEKIPSTAPLDEKIPAPVSPVVPDVEPSSSEKPAVSEYDGE-----ESTEPVHDVE 2341
QY 2319 AQQSGVPOQITLOLPPIQIQSSNAVOTQIQVWVVOAA-----SVQOLQAVQOOLRBOQQ 2373
Db 2342 TSTBEPVSDAKLPPTSPAPAISSPATVETAIPEITAPPELEKEVPEKATEQPLEKETP 2401
QY 2374 KKKQOOIEIKREHTLOASNOSEI-----IQQVYMKHNAVTEHLKOKKSMTP 2420
Db 2402 EKATEQPLEKETPEKATEQPELEKETPEKATEQPELEKETDQKATQGPESVDEKTTPEP 2461
QY 2421 AER-----EENQRMIVCNQVWKYILDKIDKEEQAAKKRRRESVEQKRSQNAATKISA 2474
Db 2462 VVKPSLDSTEDDE-----ESVESEESADKKDKNKETKEEDTDKHEEPEVPA 2508
QY 2475 L-----LKKHKEQLRAELIKKRALLDKLOLQIEVQELKRDKIKREK 2516
Db 2509 VVSEIPOPSEBAVPTTGHPLEPH--LASSTTPPAVDNR-----VGBE-----DEN 2553
QY 2517 DLMQLAQATAAACPVPY-----PVLPAPAPAPSPPPPGVOHTGLSTPTLPLVASQKR 2572
Db 2554 TTVKISSSTTTSTESPTVSAAPSTTVASQOOQPTTPPYG--HA-----PEYDEY 2603
QY 2573 KREERKSSSKKKKMIJSTSKETKDKTLVYCIQKTPYDESKFYICQDRCNNYHGRCV 2632

```

```

Db 2604 DEEEVGPCTCRVAGKLYVSAQOIPRDEDCDFCF------FRSDII 2644
QY 2633 GIIQS-----EALIDBYVCPOCQ-----STEDAMTVLTP 2662
Db 2645 CIOQSCPPPIAGCHEPIISGFCPCRYECPVMAAVALNTTSTTTSTILTP 2695

RESULT 12
ID ABB65772
ABB65772 standard; Protein; 5533 AA.
XX
AC ABB65772;
XX
DE 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 24108.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL09875.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5533 AA;
XX

Query Match 3.9%; Score 560; DB 22; Length 5533;
Best Local Similarity 19.3%; Pred. No.3.le-19;
Matches 520; Conservative 382; Mismatches 948; Indels 836; Gaps 118;

QY 325 DRSRKYWFNIRRLIEDTENENKKIWIYSTKVQLAELIDLDKDWAEELCKILEEM 384
Db 1997 DKEQKEKEIKRDLREKQREKRDREK-----ELRDKDLREKEM-REKEQR 2041
QY 385 REEIRHMDYTE-----DITNKAQSNKSFLLAANDLIELISIRAKKGIDNV 431
Db 2042 EKEIHRKQDREHREKREKQSRRAMDVEQGRGM-----RELSSYQSKRMIDAGE 2093
QY 432 KSPETEKQKNENFENDSKDAEKNREEFQSLG--KSDDKTPTDDDEQKSE----- 482

```

Db 2094 AS-----SLTALDCOHNKENAMDTTACGTPGASPTSPDNTKREKRKLSRNSP 2142
Qy 483 VGDFFSEKSNGLSESPGAGKASGSTR-----ITTRLRNPDSK-----LSOLKSOQV 530
Db 2143 VRLHRRLLSOGESNHSAGGSGCGSHQIHEDYVKRIRRMENSONIVSHSSNOQLRNBDR 2202
Qy 531 AAAAHEANKLFREGVEVLVYNSOGELSR-----LSLTKR----- 563
Db 2203 DSKREKSSS-FKEDK-----NSSSHISRPHCGGSSASSSKHHRRDKHKGASASLET 2256
Qy 564 -----EVIM-----KGNINNYFKLGO-----EGKYRVYHNOYSTNSPALNK-----HOHR 603
Db 2257 NSSIEVVVDPISQTKHNLTSSEELQSHOPKREKEREHFSHANSSSSKHKSRDHHRHHR 2316
Qy 604 -----EDHDKRRLAHKFCILTPAGEFKWNGSVGSKVLITSLRILITOLE 649
Db 2317 EKKRHSVAESTWTDEHNPQOINPHR-RISAAG-----SSSAG-----ELSSA 2359
Qy 650 NNIPSSFLHPNKAHRAHWIKAVOMCSKPREFALALATECAVKRVYMLPIWREFLGHTR 709
Db 2360 TMTSSGKLHQ--HHRRSYERKSSRGSDGHHSSSKSL-----RAK 2398
Qy 710 LHRMTSIREEK-----EKVKKKKKOEETMOQATWYKTF 747
Db 2399 LMLSSASDDTDDASKHSTFDLPDCCPNVMDKVKARSCKNMOROEKKIKAKES- 2457
Qy 748 PVKHQVMOKEGEYRVYTGYSW--WISKTHVRFVPLPGNTVYVYKSLGKTKNNMD 804
Db 2458 ----QLKOSRAKKKSTVSDGSDTEFEDRQH-----RNSSSSFHGR--YPGJLSSSD 2505
Qy 805 ENMDSXKRCRSRPPKKIKIEPDSKDEYKSDAAKADQNF-----MDISKITEKDD 859
Db 2506 DDEETHORRIS-----SDSDAHHGGGDNOCAGSTLADANEVROMQ-QN 2547
Qy 860 VKELSDSDKCKREPMVEVDMDKTESHVNCQESSQVYVYVSG-----FILRTY 912
Db 2548 LRRLLDGDGS--SED--EIRRNVMKSHHFGRKNSNSTRIASDSOSOPADPLTJKOH 2602
Qy 913 -----KKTKTSKLDGLERRIKOFTLEKORLEKIGIKGIGTSTNSKNLS 964
Db 2603 PIAPAQELAKRQGLSDEQKFKRHSNSSTIEER--KKTREI--KTLLGPFYNS 2654
Qy 965 ESPVITKAKEGCQSDSMROESPANNDOPEDLIGCQSDSS-----VLRMSDPS--- 1015
Db 2655 ETTYNGKLKEVSPETRKHKHKKRR-----LKSSSTADTSAQTPLYMPLTPISLP 2706
Qy 1016 -HTTKLVPKQVLDVSL-----RSPETKCKQKSIENDIEKXSDLAS 1059
Db 2707 VHSSECKTEFDNFDLKTCESSIFLEISAGERRHKKERKKRKLARMTATVPN-- 2763
Qy 1060 RGOEPTKSKTKGNDFIDDSKLASADDIGTLCKNKKPLIQEESDTIVYSSKSALHSSVP 1119
Db 2764 ---SFTTNDT-----SSEKLSKEERHRLKSKKSKSMDSCHTKIYNSGA-H---P 2808
Qy 1120 KSTNDRATPLSR-----AMDFEKLGCDSBSNSTLSSSDTVSIQDSEEDM 1167
Db 2809 STSPSLPATPTSPASTAQTSKRGEDKMEFIFGIISDEESQPEQAETNKDIIIPSS--- 2864
Qy 1168 IVQNSNESISEGFRTRREDVLEPLKCELVSGESTGCNEDLPRKGTENANK-----KPS 1223
Db 2865 -VSTTGPVSAALQTYKO---EP-----STPNS---KNEAHNLQVTHEPE 2903
Qy 1224 OOKLEERPVNKCSDQIKLTKTTDKNNENRESEK-----KQORTSTFOJNGKNDPKYIL 1279
Db 2904 QOQOQLER---SRLSGGSSSSSHADREHRRERKREKKREKSEKREGOQNOIHOKSSK- 2955
Qy 1280 KGECLKELSESXVSVGNEPVKNVINKIIPENDIKSLTVKBSAIPFINGDIVIMEDFNER 1339
Db 2956 -----VETKXDDNSVMDDEAGRALBAQ-----LMSDFDK 2986
Qy 1340 NSSETKSHLLSSDAEGNVRBSLETLPSTK---ESDSQTOTTPASACPEPSNSVNOV--- 1392
Db 2987 PISE-----EATPSTAAVYRSDMTDVFRESDN-EDNNSVDMTKOGV 3026

Qy 1393 --EDMEIETSEVKVYTSSEITSEESNLSNDEINGLPIKNENYNGSKRKVITEVT 1450
Db 3027 KSEOEONHKSJKDKKKKKRSEKOEKLL-----QOQRRESLPNVA 3067
Qy 1451 TMTSTVATESKIVIVYEKDKOT-----VVSINECAKSTV-----TTTTTV 1493
Db 3068 STSSAPPTPGKLTVVVQAASKHADQLQAKHISPSVPCPSPSLPCLJGDDDDALHTPK 3127
Qy 1494 TLTSTPTGSGVDIISVKROKTV---VTVTVTSLT-----TYGGLYMS--- 1536
Db 3128 AKPTTPSSKNGDLTPSRKPKRLISPETPIANSSTLSTOSAEPTVSSGVISSSALA 3187
Qy 1537 -----MTVSKEYST 1545
Db 3188 TPTSTTAAGVSAARGLDNPSTASAOCKKSFIPGFGQLDDRISESAGVISAEFRS 3247
Qy 1546 -----RDKYVLMKFSRPKKTRSGTALPSYKKEVTKSKK----- 1579
Db 3248 TSLDNIADPEKIPYASPRATKPLDKLEESKSRVTISOETESAVSALLGSEFSTST 3307
Qy 1580 -----SIFVLPNDDLKLARK-----GGIREVFPYNNAKRAL 1612
Db 3308 DYSLDGMDMSVSNLEPTTLVIAEPDEBALAALAIETAGRPASTILEEP---EKERER 3363
Qy 1613 DIMPYRSP-----RPTGKITWRYL-OTVKSLAGVSLMLRLYMASLRMDMAKVPFG 1665
Db 3364 EAPRPDPRAELIESRVEVLELREBELKNVQSLKHEDNM-----DIKADTPQSE 3412
Qy 1666 GSTRTETSETLTETTLIKRQVGRVGIREFEYCIKKIICIPGIVETPKETPPRQKGLRS 1725
Db 3413 RDLQIDT-DTEENPDE--ADSSGP-----SLKIDETVSSSSPEKSISSNN 3454
Qy 1726 SALRPKRPEPRKQGVLIETVWAELELELWEIAPAEVEKEKAQVQAOKKRLQOK 1785
Db 3455 S-----PTEPATNIDI-----PNVESQPKLSNESTQPSVITKL 3489
Qy 1786 P-----TVIATSTSPSTSTTSPISPAQVNAVARSV---TGTGMVLTVTVGSPA 1835
Db 3490 PFLDLPKTVPALPSPVYKIEPTLSKIQQLVQVQVTLPAHSTGSGISANSVINDL 3549
Qy 1836 TVTFQONKNFHOTFATVWKQOGSNGVVOVQOKVLGIPTSGTSGOQFTSPQ-RTATV 1894
Db 3550 NVWISSCSNTSASATASASISISFGSPASQNAM---FOASTPQKQPTTPQOALRTSL 3606
Qy 1895 TIRPMTSGSGGTTN---SOVITGPOI---RPG--MVIYRPLQOSTL---GKALIIR- 1941
Db 3607 IMOPRTISITPEQTPHFAVQWLSPOSHHPQOPGTVMGICIRAPSRPHSLSPGKVAQSR 3666
Qy 1942 -----PYMVOPGARQOVMTOIIRGQPVSTAVSAPNTVSSTPGOKSLTSATSTNI 1991
Db 3667 LVGOLSPVGRNVPNSQ-SPQO---QVOQTOQOONALITSPOSSNISPIASPTTVLSSSUS 3722
Qy 1992 QSSA---SOPRRQOGQVKLMAOLQTLQGHGNOGLTVLTIQOGQGTGQDLQILPQCVT 2048
Db 3723 PTTSKVYNSYQRPNOQ-----VPOQSPRKSVAEVOGTPTOLMT 3758
Qy 2049 VLRPGQOLMAQAMPNGVQRFPLRTATATATTTTUVSTTAAGSEQOQSLTSPQM 2108
Db 3759 I---PQOKTPRIQVHNPT---TISKVUTVQROATQSOVASSPRGS-----LPRK 3805
Qy 2109 QVNOCKTLRPAOSSVSGRAKAPQT-AOPSAFROFQOPQSAFREPOTREVOGTIVS 2167
Db 3806 NVH-----LNAHON-----QOQOVIYAKTANOQNOQNOQNOQNOQNOQNOQNOQ 3849
Qy 2168 SHVPSDAQPTAAOSSKPOVLAOSQPOSUNVQOSRVRVQVSPQTRIPRSTPSQLSFGQ-OS 2226
Db 3850 -----QOOLHGOQO--OITSARQONHOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3901
Qy 2227 QVQTTTSQPIRQVHTSIQILPSQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2283
Db 3902 QAOQOQFNQ--QIOQHOQOQO---OHQVOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3952


```

Db 2809 STSPSPAPPTSAPTAQTAKSKGECMEFIIGIISDEESQPEQAEYTKNDIIPSS----- 2864
QY 1168 IVQNSNESISEOFRTREODVEVLEPLKCELVSGESTGNCEDRLPVGTANGK-----KPS 1223
Db 2865 -VSTGPRIVSALQIYKQ-----EP-----STPN-----KNEAHILQIYHEPB 2903
QY 1224 QOKKLEERVNKCSDQIKLKNTTDKNNENRESEK-----KGQRTSFFQINGKDKPKIYL 1279
Db 2904 QOQOOLER-----SRLSGSSSSSHADREHREKREKRRKREKQOQOQIHOQSK----- 2955
QY 1280 KGCCLKEISESRVSGNVKPYNNIKIIPENDIKSLYKESAIRFINGDIYMEQNER 1339
Db 2956 -----VETKDDONSVMDEGRALEAO-----LMSDFIK 2986
QY 1340 NSEETKSHLLSSDAGNRDSELETLPSTK-----ESDSTQTTTPSAPCSNSVNOV----- 1392
Db 2987 PISE-----EATPSTAATYRSMDTVFRSDN-EDNNSVDMTKQCV 3026
QY 1393 --EDMEIETSEVKKYVTSPTTSEESNLSNDFIDENGLRPNKNENNGESKRKVTITEVT 1450
Db 3027 KSEQOQHKSKDKKKKKKSKSEKQKLL-----QOQRESLPNVA 3067
QY 1451 TMTSTVATSKTYIKYKQDKQT-----VVSSTENCAKSTY-----TTTTTV 1493
Db 3068 STSSAPPTGKLTVNOQAASKHADLOLDAKHISSPVCKPSPSLPCLISDDDDDLATPK 3127
QY 1494 TRLSTPSTGSDVDSIKVEKQSKTV-----VTTVTYDGLT-----TTGGLVTS----- 1536
Db 3128 AKPTPSSSGNGGLTPSRREKPRILSPIRKPTIANSSTLSTQSAETPVSSGVISSALA 3187
QY 1537 -----MTVSKREYST 1545
Db 3188 TPTSTSTAAGVSAAPGLDNPSTASAOCKKESFIPGQGLDDRISESVAOSISAEFNS 3247
QY 1546 -----RDKVKLMKRSRKRTSGTALPSYKRFVTKSKK----- 1579
Db 3248 TSLDNIADPKIPVAPSPRAKPLDLKLEBSKRVITISOETESAVSALIGSEGTSSPT 3307
QY 1580 -----SIFVLPNDLKLARK-----GJIREVYFENYNAKPAL 1612
Db 3308 DYSLDGMDMSVNELETPTLVIAPDEDEALAKAIETAGEPASLLEP-----EMEPEK 3363
QY 1613 DTPVTPSP-----RPTFGITWRYRL-QTVKSLAGVSLMLRLIMASLRMDMAAKVPPGG 1665
Db 3364 EAEPPDEPAEIESEPVVLELPEELINKAVOSLKHEDMM-----DIKADPQSE 3412
QY 1666 GSRRTSETELITTTILIKRDVPGYGFRCIRKILCPICGVPETPKETPTQKGLRS 1725
Db 3413 RDLQIDT-DTEENPDE--ADSSGP-----SLKIDETVOSSSPEKSIINN 3454
QY 1726 SALRPKRPETPKQTFVILETWAELELWEIRAFERKEKEKAQAVQOAKKRLQOK 1785
Db 3455 S-----TPRETANIDI-----PNVESPKLSNESTPOPSVITKL 3489
QY 1786 P-----TVIATSTSPSTSTSTISPAOKVMVAPISGV-----TTGTVMVLTWKVSPA 1835
Db 3490 PELDTPKTVAGLPSPVXIEPRTISKLOQPLVQVQTLAPRHSSTGISAISAVINL 3549
QY 1836 TVTFQONKFNQTFATWVQOGOSNGVVOQKVLGIITSSGTSGQTTSTQFQF-RTAIV 1894
Db 3550 SNISSCSNTSASATASASASISFCSPTASQAM--POASTPKQGITPQOALITQSL 3606
QY 1895 TIRPNTSGSGTTSN--SOVITGPQI--RPG--MTVIRPPLQOSTL--GKAIRF- 1941
Db 3607 INQPTLISPEQTPHNAVWLSPQSHHPQPGTYMVGITIRAPSRHPSLHSPRGVAGQR 3666
QY 1942 -----PYWVOPGAPQOVNTQIIRGQPVSTAVSADNVVSTFGQKSLFSAITSTSN 1991
Db 3667 LVGOLSPVGRPMVNSP-SPOQ--QVOQTOQOQHALLITSSPOSSNISPLASPTTFLVLSNS 3722
QY 1992 OSSA--SQPRPQOGQVLTMAQLQLQGHGNGGLVNVIGQGOQTGQGLIQGVT 2048
Db 3723 PTTSKVNSIQPRMO-----VPOQPSPKVAEVOYTPQLMT 3758

```

```

QY 2049 VLRPGQOLMQAMPNGTVORFLFRLATATATSTTTTSTTAAGTEBOROSKLSPM 2108
Db 3759 I--PLQKMTPIQVUNHNP--ILSKVTVVQROAQNSQVASSPRLGS-----LPRK 3805
QY 2109 ONHOKTLPRAOSSSVGRKAKROPQ-ADPSARPOROTOPQSPRAQVQTPREVQOTTVAS 2167
Db 3806 NVH--LANOH-----QOQPOVIAKMTAHQHOQMHQOQFMHQMIQROQHNO----- 3849
QY 2168 SNHPSAOPRTHQSSKROYAAGQPOPSNVGOSRPVVOVSPQTRIRPSPOLSPOQ-OS 2226
Db 3850 -----QOOLHGSO--QITSAPROHMOHOHOQOQOHHNHOQNLHQO.LHAQHPYQKH 3901
QY 2227 QVQTTTSOPRIQPIRHTSLQIPRSQGPQSPQVQVQV--STQTLSSGQTLNVQVSSPSRPQ 2283
Db 3902 QAGQQRNQ--QIQOHQSQ--QHVOQOQNAQOQHLSQOQHOSQOQNLQ--QHQAQ 3952
QY 2284 LQIQOPQ-QVIAVPOLOQO-----VOVLSQIOSOVVAQIOAQOSGVPQOIK 2329
Db 3953 QOIQOIQOKIQOMHGPQOQKSPQGVGHLGSTIFASQOHNQOLPAR-----GVPOQ-- 4004
QY 2330 LQPIQIQOSSAVQTHQIQNV-VTQVAAVQEDQORV--QQLMDQOKKKKQOIELKREH 2386
Db 4005 -QHPOQLSHSSPCKPNTLVSNQGVQPRPI--LTRVGSHPQNOQOQLPHQOSSGHPH 4060
QY 2387 TLQASNOSEIIOKO--VVMKHNATV--EHLKOKKSMTPAREEENQRMIVCNOVKYIL 2440
Db 4061 QKQLSSPGANLPLQGTPLNYIQNTPKIIVQGHVIAQNVQPRPQVGNAINHPQNG-KKDSF 4119
QY 2441 DKIDKEKQAAKKRKRESEVQKRSKQNTKLSAL-----LFKHEDLRAELTKKRAL 2493
Db 4120 PCHNVETPTPAMSQAKTSESVYIRPTPTTGLAVISANTVSGSLTPEENLIKISQPKODEL 4179
QY 2494 LQKDLQIEVO-----EELKDLKIKK-----EKDMLQAOATAVAAPCPVTPPLPAR 2541
Db 4180 IEQDSK-EVDSQYWSKKEVNIIDSVIKLDTPLASKAKKAVEQAL----- 4224
QY 2542 PAPPSPPPPPGVQHTGLSTPLPYASQKRKREKEDSSSKSKKKKK 2589
Db 4225 -APAPIPNQPGNQ--SMAQETALPTTSMVSNVNSNDHDEDETEHQ 4269

RESULT 14
ABB66811
ID ABB66811 standard; Protein; 6815 AA.
XX
AC ABB66811;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polyptide seq ID NO 27225.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX PN MO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX DR MPI: 2001-656860/75.
XX DR N-PSDB; ABL10914.
XX

```

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 27225; 21bp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence 6815 AA:

Query Match 3.8%; Score 549.5; DB 22; Length 6815;
Best Local Similarity 17.7%; Pred. No. 1.4e-18;
Matches 611; Conservative 346; Mismatches 1236; Indels 1063; Gaps 146;

QY 2 VSEEEEDDADAETODSEDEDEDE---MEEDDDSDYPEEMDDDDASYCTESSFRS 57
DB 1436 IDNOKTKKISEPPTSESDTIEEYPRKISEHDEDLQTOEYSDVDSL-----P 1486
QY 58 HSTYSSTPGRRKPRVHPR-----PILEP-----KDIPLEPKSSSEDLMP 100
DB 1487 KSKRKSSTKQKESLPGISLYTIRIETTPETTEKTYEDGKVVAVINKRIRKKAKAP 1546
QY 101 NEHMANIYAIY-----VLRNFGVLRLSPFRFD-----FCAA 134
DB 1547 KEYIIEVTEYEDNNPDAVDVIRIETTPSIDSKPDQDHKIQVQEKPKTESIDNTYQK 1606
QY 135 LVSOE-----QCTLMAE-----MHVLLKAVLREEDTS 162
DB 1607 LIDDEIPQVYDHKEFKATVLETSPESKAKKIKKHHKKTTEVIGDIPITVEVITQETED 1666
QY 163 NTTGPRAD--LKD-----SYNSTLYFTDGWTWPEVLAAYGDSKEVHNVL 205
DB 1667 DEDKPRPEVTLKETDENAEAPKEQVEITETKALDEIPSEVTVAEITDQPE--EVL 1724
QY 206 PYQEAEDYPRGYENK-----IKVLOFLVD--QFLTTNIAREELMSGVIOYD 251
DB 1725 PAQCKK-----PIKQKKLKPEDVNTVYVLELETPETP-----ETIPEDA---D 1768
QY 252 DHCYVCHKLGDLLCCETCSAVYHLECYKPRLEEVPEDEMOECVANKVGVYDCAVEIQ 311
DB 1769 DK-----POVIEDISENVQVQIIEEDGTPK---QVEIK 1800
QY 312 KNRPIYHEPIGYRSRRKWF-----LNRRLIIEEDEN- 346
DB 1801 KKVYSPKHP-----KEOVFEITETRPSPDEPLAEVTVLEIEGKNDIVIPQEKTV 1853
QY 347 ENEKIYYSTKYQALBILDCLDYWEALCK--ILEMREBEIH-----RHMDI 394
DB 1854 KKKPKLPEDIQSYVIVLEEFENPQWPASTEKRIIDIASIEIIVPTEEDGTYKKEVEV 1913
QY 395 TEDIITNARKGSNK-----SFLAANEELIESIRA 423
DB 1914 KKKVYSRKQGTKNQVFEIETKTSDEPLAEVTVILELSGDSQSEVTVILPKKKPKTKTKIKL 1973
QY 424 KKGIDID-----NKSPEETEKDNKTENDSKSDAKNREPEFQSLKSDSD 469
DB 1974 KPDEVESTVYVNVLEEFCEPQSFESPEPTGABHETTKTKPKPKPIYAKABENWIL--IEE 2031
QY 470 KTPDDVEQGSKEVY--DFKSEKSNGLSESPGAGKSGSTRITRLRNPSDKLSQLSK 528
DB 2032 MAPETVLENIYNEIGEYKQVTKTKLKKKG-----PREYILEIKE- 2073
QY 529 OVAANAHBANK-----LFKEG-----KEYLVVNSGCEISRLSTKKEVKMKN 570

DB 2074 -----TYEENKREGDIEITTEELVEGSPDASDDPVIYVQ-----KIKKKRPVXD 2120
QY 571 INNFK--LGQEKYRVYVNOYSTNSFALNKHQHEDHDKRRHLAHKCLTPAGEKNG 628
DB 2121 LDKYIQQLIEQ-----IKTPLEEYEPTEMDSKK-----PKKVK--- 2157
QY 629 SVHGSVYL-TISTRLRLTITOLENNIPSSFLHNPMASHRANWIKAVOMSKPREFALAI 687
DB 2158 -SHNKKTIEVIGLPLVTI-----HEEN---VEDIVSEPDMEPTKTL 2195
QY 688 LECAYKPVVLP-----IKREP-LGHTRLHMTSTIEREKEKVKKKKEQK- 732
DB 2196 LD-EIKEIPQLPDSSSKYLVNIDSDEFGADKPIQTOOPIKKEPKLKKKDVVEPVSL 2254
QY 733 -----EEETMOQATWVKTPEPVKHQVWVKQKEEYRVYGGGWSISKTHYREVP 782
DB 2255 EAFDHTVKVYSEPTLEGT--VKEVYKKRVSKRG-----SDHIFEIT- 2297
QY 783 KLPGNTNVYRKSLGCT--KNNDENMDESDK---RKCSRSPPKIKIPEPDE----- 829
DB 2298 ---ETTSEDRPAEVTYVLELSDEVLDSEKPKHERKLVKKPKOLKDDVEEYIINITE 2353
QY 830 -----KDEYVGSDAKGAQONEMDISKI-TEKKDQ-----VKELDSDSKP 871
DB 2354 EFTQPIPVGLVEDEVEKVKQETKPKKSPITYIATEQEDNNNNYDALKEDLDPIERA 2413
QY 872 CKREPMVEDDMKTESHVACOESSQVDVYVNSGFLRTSYKKKTKSKSLDGLLEARI-- 929
DB 2414 LEKPSPLEYTTISVEDSVEGEOKOPKPKKISKPKIKQPSVDKSDYLWVNISESTID 2473
QY 930 ---KQFTL---EERQ-----RLERIKLEGKIGKIGTSTNSKNLSSEPVITAK-- 973
DB 2474 EPIPEYVYVTEAAEKRPSEPTPKVELETE---AVEKEVTDGDKGTTQKOSVTRKIK 2529
QY 974 -----EGCOSDSMRQEOFPANNQO-----PEDLIQ 999
DB 2530 KLVGPREELIEIYETGTGTPEYEVITVTEVOEKSEKPEEKAKTVAKKAKIPKDDIQ 2589
QY 1000 GCSQSDSVLRMSDPHTTKLIPKRDVLVDVSTRSPETKCPKONS-----IEND 1049
DB 2590 DYTQK---LIBODIKTELEKEYEKIDDEPVAKMKRPIKKYKQSEQKETEHEPIEDK 2645
QY 1050 IEKVSDDLASRQK-----PTKSKTKGDNFFIDSKLAS---ADDIGTL---I 1091
DB 2646 PVKEISEYSDDEDEKPLVAAVEFIPKEPPEEPFIVVLEEVESKRPDEGKVRKY 2705
QY 1092 CKNKKPL-----IOESDITVSSSKALSHSVKSTNDRATPLSR----- 1132
DB 2706 VKTKIKQNRGSEVYVHDIVIEIDTDPNESAIVTTVTTPETPDQOPSVYQKRTKIKK 2765
QY 1133 --AMDEFGKLGCDSESNSTLENSDVTYSIOD-----SSE--EDMIVONSNEISQFRTR 1183
DB 2766 DEVEDFVKRY--IEEAPQEGSVLDLVIEDVPKRSSEKRRKKPKDKDHTSVEE--TPH 2822
QY 1184 EOVYVLE-----PLKCELVGSESTGNCEDRLPVYGTEAN-----GKPSQOKK--- 1227
DB 2823 EDEVLITIESVPEDSPISDDLIT-----VDSVPILIEEPENKVNQIEDTKPKPKKPRP 2876
QY 1228 -----LEERPVKCSDOJK---LKNUTDKNNENESK 1258
DB 2877 SAKILEBNPDETVKRPLEALHTDSLEKPDVOEFSISIKEEQKITHHEKKSSKISSE 2936
QY 1259 KGORSTY--FOINGKDNKPKIYLGCECLKEISRSRV--GNVBPVNNINKIIPENDIKS 1315
DB 2937 QKQOPSTEOYEIVTEHD---LKPEEKRFYQVYQISEFNVEETKDDTGKV-----HKQ 2987
QY 1316 LTVKESAIRPFTNGDVIMEDFNERNSSETKSHLLSSDAAGNTRDLEFLPTSKESDSIQ 1375
DB 2988 VTKRMLRPRAGGEIIEIVVRDDQPEAEITTVIEVEPEPVN--QDKRPKPKK---TR 3042
QY 1376 TTTPSASCPESNVNOYEDMEIETSEVKKVTSSPISSESNLSNFIDENGPIPKNNEN 1435


```

Db 3043 KYKKDDIHVIOKLIETPKTELEKIEPEPIYKOKPLDSPIDVLDSEKPEVOKK 3102
Oy 1436 VNGESK---RKVTIETVMTSTVATESK-----TVIKVEKDKOTVVSSTENCA 1482
Db 3103 KSKSTVPEHEETPVQOYAKVNVVEEAEQPEIPVQOILEVKEVDEVAEVI---TED-G 3158
Oy 1483 KSVVTTTTTVTKLSPSGSVSDIISVEQOSTVTTVTBLSLTGGTLVTSMVSKS 1542
Db 3159 KPVQEKTKRVKIKKIGPEBQTEFKITMIESEDNSVTVIWDEE-----PEIASPOSIEE 3212
Oy 1543 YSTRDVKLMKSFSPKTKSGTALPSYRKFWTKSTKKSIFVLJPNDDLKLARKGIREVP 1602
Db 3213 HEQOSKEKLA--PKPKTYRRAKKDDLSYVKKLIEEL---PKVLEXY-EKVEPEKXP 3266
Oy 1603 YENYAKPALDIMPYSPSPRTGCI---TWRYRLQTVKSLAGVSLMLRLMALSRLMDMA 1658
Db 3267 -VKLTVSDSIPREPPKPKDSQPSISVLBDTKPKKTKTPK-----TPKREDND 3311
Oy 1659 AKVPPGGSTRETSE-TEITTEITIKRDVGPYGR-----FENCIRKILCP 1705
Db 3312 QOVPDEPTETVDTNDIPELTPTQTAQPEDTATQITPSAQEEKSTODDTKDTIQKTV-- 3369
Oy 1706 IGVPEPTEPTPORKGLSSALRPKRPETPKQTPGVIIETWVAEEL--ELMEIRAFAE 1763
Db 3370 -----KHKTTPDQKSVETSEL-----PEVHKDYOISIIHEELVEEPEKILLEVVIDE 3420
Oy 1764 RVEKEKAQAEQOAKKRLDQKPTVATSTPSTSTSTISPAQ-----KVNV----- 1812
Db 3421 VAEVESQPIVEEVEDEEPOPAETEVEDVTPKSKKKKVVKKKTDHDLIKKMLEQEI 3480
Oy 1813 -----APISGSVTGKMLTTLTVGSPATVTFQONKN 1844
Db 3481 EKTELEKYEKIEFDVPKKLPEFALEPIKIERKQPKTKVILTDADVPKVKLPSKR 3540
Oy 1845 FQOTFATWKOQSGNSGVVOYQOKVLGIPSSGTSGQTFSTQPRATVVTIRPNTSGSG 1904
Db 3541 -----KEKPAEELTQ-----LPKRRLKRAMLVETRP---APLIPKTTIDIG 3579
Oy 1905 GTTNSQVITG-----POLRGMVTIATPLQOSTLCK-----AII RTPV 1943
Db 3580 AIKDNGLSRNIEAEELKFKPHKTKIKIKIDLEKVELEKYEKVISSSEPEKTPY 3639
Oy 1944 MQOPGAR-----QOYMTQIINGQPVSTAVSAPNTVSSTPGOKSLTSANSTSIQSSAQ 1997
Db 3640 KRPEKAPKEEKOEDVKLIGKKKKKEEAPENV-----TLKNI----- 3680
Oy 1998 PRPPOGGVUKLMAOLTQTOGHGNGGLVVIQGGQTTGOLQLP-----QGVTVLP 2051
Db 3681 POKPQVEEVEELKOKPKEVE-----IVEEQTKPKDGEVVEPEPSEFDRPEYVP 3732
Oy 2052 GPGQOLMAAMPNGTVORFLPTLATTTASTTTTSTVSTAAGTEGORSKLSPOMYH 2111
Db 3733 DELEQLEHPEIDE-KVKKPSKTKYKPKDKSKSPETIVEIAGVPEKEEA--IPQDQV 3789
Oy 2112 QDKTLPPAOSSSVGPAKKAPQTAQPSARQOTOPQSPAQPEVQTOPEVQTOPTVSSHVP 2171
Db 3790 FRKPERDAPEEDSEIKLRP-VPOASKDENPDEQALVTPEKAEPIRQIEDIAIDDEKRP 3848
Oy 2172 SEAOPTHAOSKPOVAAQOSQPSNVQO-----SPVAVQSSQRIHRPSPQSLSPOQ 2224
Db 3849 KSKPKKQVOPKQOELAKEPEEVEVSKKEEALVDBKPIETEKPKDVKEKKPKE--AP 3905
Oy 2225 OSQOVOTTTSQP-----IPLOPH--TSLOIPSGQPOPOPOV-----QSTOTLSSG 2268
Db 3906 VSEVVVIEEPEPEVEPEIPEYKITTIVLEBEDAKRHQYKVIDFDERQETTEVEIEE 3965
Oy 2269 QTLNOVSVSSPSRP---QIQIQOPQPOVAVPOLQOQOVVLISQIOSQVVAQIQOQSGVP 2325
Db 3966 KVVTKRKKRKKPOQPEFEVTLKEPKEE---QIQPQV-----VSAEISLPIEPEQO-KP 4014
Oy 2326 QOQIKLOLP-----QIQQSSAVQTHQIQNV-----VTYQAASVQEOLOLR-VQOL 2368
Db 4015 EOYEVELKITQTTPEEPNDVQIAVKEKVTKPKVKKEDKIYVVEAEKEKOPVEITIV 4074

```

```

Oy 2369 RQOQARKKO-----QOIEIK-----REHTLO 2389
Db 4075 EKQEEKKSEKPSYEFKISQISQIEKPLEVAEAPPEETPKVVEKVAEKDSYEFLLK 4134
Oy 2390 ASNOSEIT-----OKOVYKHNVAIEH-----LKQKKS 2417
Db 4135 ETDKEKVIIVDDQPEEBAVEVFKKRPKEPAVEAEFVTEPKIYEENSVEITAIKOKKT 4194
Oy 2418 MPFAREENORMI-----VCNOVM-----KYLIDKIDKEEK-----QAKKR 2454
Db 4195 KPKKDEEBAQLAIKVESEAPVAEEVSEAPSKTIVEEVAIEEKPKEFTIRVSESEK 4254
Oy 2455 KREESVEQ---KRSQONATKLS--ALLFKHQBQRAEIL-----KKRALD---K 2496
Db 4255 PEESVTEQFTVKRKRSVTFADBPATEIYIKESKPAVEVTEDAHIKTKPKKVTQVEAE 4314
Oy 2497 DLQIEVOEELKRDLIKKEKDMQLOAATAVAAP-----CPPTVPLPA----- 2540
Db 4315 ELKIKITEEVPOEIPILEVSEBEVITETKTAAPVVEEKTYKIGIKETEPEKPAEAIVEE 4374
Oy 2541 --PPAPP--SPPPPGVOH--TGLLTP-----TLVVASOKKRREBEKD----- 2579
Db 4375 EEPVTEPELEAPKRPVEFEHKVRYIEETPRELVEEYIEEYKVIIRKKAPKEPIKEEPA 4434
Oy 2580 --SSSKSKRKMISTTS-----KETKDKTKLYCI-CKTPYDESKFYIGCDRCQMW 2626
Db 4435 EYVSTPKPVEEVEATSSIAVIEQPTBEEADKITIIEEETPQE----- 4481
Oy 2627 YHGRCVGIHQ---SEALDIEYVCPOCOSTEDAMTVTLPTLEK--DYEGIKRVLRISQIAK 2682
Db 4482 -----LVQZIEIEIIEVEEPKAPBEPDND--FTFATKQSEKRPVEELPPEBOVTIOKKR 4532
Oy 2683 MAMPFLEVPDNDAPDYGVYIKEPMDLATMEERVOR 2718
Db 4533 KKAPVEVVEEPEA-EFLVKPKTIPQOEVTEBAKITK 4567

RESULT 15
ABB64877
ID ABB64877 standard; Protein; 3080 AA.
XX
AC ABB64877;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 21423.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-65660/75.
XX
DR N-PSDB; ABL08980.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:26:07 ; Search time 323.179 Seconds

(without alignments)
5548.021 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MVSSEEEEDGDAEETDSE.....KLKGFKSRSHNNLSTAS 2781

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_Main.*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14333	100.0	2781	US-09-698-295-10	Sequence 10, Appl
2	14170	98.9	2907	US-09-698-295-1	Sequence 1, Appl
3	13428	93.7	2759	PCT-US02-21179-15	Sequence 15, Appl
4	3363	23.5	845	PCT-US01-14827-16042	Sequence 16042, A
5	2695.5	18.8	557	US-09-629-14969-14273	Sequence 14273, A
6	1557.5	10.9	976	US-09-614-150-32997	Sequence 32997, A

7	1557.5	10.9	976	US-60-191-637-32572	Sequence 32572, A
8	1557.5	10.9	976	US-60-191-661-25768	Sequence 25768, A
9	1493	10.4	786	US-60-173-464-22080	Sequence 22080, A
10	1478	10.3	273	US-10-029-386-32957	Sequence 32957, A
11	1474	10.3	328	PCT-US00-05989-816	Sequence 816, App
12	1474	10.3	328	US-09-925-297-816	Sequence 816, App
13	1368.5	9.5	412	US-60-389-967-1080	Sequence 1080, Ap
14	1368.5	9.5	412	US-60-412-418-1080	Sequence 1080, Ap
15	1152.5	8.0	246	US-09-758-472-6374	Sequence 6374, Ap
16	1152.5	8.0	246	US-10-235-926-6374	Sequence 6374, Ap
17	982.5	6.9	238	PCT-US98-17709-69	Sequence 69, Appl
18	982.5	6.9	238	US-09-729-835-80	Sequence 80, Appl
19	981	6.8	237	US-10-220-120-243	Sequence 243, Appl
20	941	6.6	237	US-09-758-472-8851	Sequence 8851, Ap
21	941	6.6	237	US-10-235-926-8851	Sequence 8851, Ap
22	886	6.2	217	US-09-758-471-3462	Sequence 3462, Ap
23	886	6.2	217	US-10-235-953-3462	Sequence 3462, Ap
24	871.5	6.1	1937	US-60-150-584-1021	Sequence 1021, Ap
25	781	5.4	246	PCT-US01-08656-8915	Sequence 8915, Ap
26	687	4.8	212	US-09-758-471-3112	Sequence 3112, Ap
27	687	4.8	212	US-10-235-953-3112	Sequence 3112, Ap
28	672.5	4.7	997	US-09-614-150-15453	Sequence 15453, A
29	672.5	4.7	997	US-60-173-464-19825	Sequence 19825, A
30	672.5	4.7	997	US-60-191-661-12281	Sequence 12281, A
31	641	4.5	272	US-09-270-767-31963	Sequence 31963, A
32	641	4.5	272	US-09-270-767-31963	Sequence 31963, A
33	641	4.5	272	US-09-270-767-31963	Sequence 31963, A
34	576.5	4.0	4007	US-60-173-464-19825	Sequence 19825, A
35	572.5	4.0	2768	US-09-614-150-31983	Sequence 31983, A
36	572.5	4.0	2768	US-60-173-464-24710	Sequence 24710, A
37	572.5	4.0	2768	US-60-191-661-15155	Sequence 15155, A
38	572.5	4.0	2768	US-60-191-661-25036	Sequence 25036, A
39	560	3.9	5533	US-09-614-150-24108	Sequence 24108, A
40	560	3.9	5533	US-09-614-150-40272	Sequence 40272, A
41	549.5	3.8	6358	US-60-191-661-27445	Sequence 27445, A
42	549.5	3.8	6358	US-60-191-661-22054	Sequence 22054, A
43	549.5	3.8	6815	US-09-614-150-27225	Sequence 27225, A
44	536.5	3.7	3059	US-60-167-217-21631	Sequence 21631, A
45	536.5	3.7	3059	US-60-171-627-1719	Sequence 1719, Ap

ALIGNMENTS

RESULT 1
US-09-698-295-10
; Sequence 10, Application US/09698295
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-698-295-10

Query Match 100.0%; Score 14333; DB 20; Length 2781;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSSEEEEDGDAEETDSEDEDEDEDDDDDDSYPEEMEDDDDDASYCTESSFRSHST 60
DB 1 MVSSEEEEDGDAEETDSEDEDEDEDDDDDDSYPEEMEDDDDDASYCTESSFRSHST 60

Qy	61	YSSPHGRKRVRHPRSPSPLLEEKDIPPLPEFPKSSSDLMWPNHNNVTAIEVLRNCEV	120
Dp	61	YSSPHGRKRVRHPRSPSPLLEEKDIPPLPEFPKSSSDLMWPNHNNVTAIEVLRNCEV	120
Qy	121	LRLSPEFPEFCAALVSQEBQCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY	180
Dp	121	LRLSPEFPEFCAALVSQEBQCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY	180
Qy	181	FTDGMTPEVLYRVYCESDKYEHNVHLYPYQBAEDY PYGVENKIKVLOFTLVDFLTNNARE	240
Dp	181	FTDGMTPEVLYRVYCESDKYEHNVHLYPYQBAEDY PYGVENKIKVLOFTLVDFLTNNARE	240
Qy	241	ELMSEGVIOYDDHCHRCNHLGDLCCGECOSAYVHLECVPRLEEVPEDEMOCCEVCVAHV	300
Dp	241	ELMSEGVIOYDDHCHRCNHLGDLCCGECOSAYVHLECVPRLEEVPEDEMOCCEVCVAHV	300
Qy	301	PGVYDCAEIOCKKMPYIRHEPIGYDSRRKYVFLNRLIIEEDTENENKKIKWYSTVQ	360
Dp	301	PGVYDCAEIOCKKMPYIRHEPIGYDSRRKYVFLNRLIIEEDTENENKKIKWYSTVQ	360
Qy	361	LAELIDCLDKOYEAELCKILEEEMBEIHRHMDITEDLTNKRGSNKSFLAANEELIES	420
Dp	361	LAELIDCLDKOYEAELCKILEEEMBEIHRHMDITEDLTNKRGSNKSFLAANEELIES	420
Qy	421	IRAKKGIDNVKSPEEHEKXNETENDSDAKENRPEFDOSLEKDSODKTPDDPEQK	480
Dp	421	IRAKKGIDNVKSPEEHEKXNETENDSDAKENRPEFDOSLEKDSODKTPDDPEQK	480
Qy	481	SEVDFSEKSNGBELSEPSGAGSGSTRITRLNPNPSKLSQKSOOVAAAAHEANKL	540
Dp	481	SEVDFSEKSNGBELSEPSGAGSGSTRITRLNPNPSKLSQKSOOVAAAAHEANKL	540
Qy	541	FKEBEVLYVNSOGELISRLSTKKEVIMKGINNYFKLGOEGKYRYHNOYSTNSFALNKH	600
Dp	541	FKEBEVLYVNSOGELISRLSTKKEVIMKGINNYFKLGOEGKYRYHNOYSTNSFALNKH	600
Qy	601	QHRDHDKRRLHLAKFCLTPAGEFKNNGSVHSGKVLITLTLLTTOLENNIPSSFHLPN	660
Dp	601	QHRDHDKRRLHLAKFCLTPAGEFKNNGSVHSGKVLITLTLLTTOLENNIPSSFHLPN	660
Qy	661	WASHRAWIKAVQMSKPREPALALAILCEAVPYVMLRIMREPIGHTFLHMTSIEREE	720
Dp	661	WASHRAWIKAVQMSKPREPALALAILCEAVPYVMLRIMREPIGHTFLHMTSIEREE	720
Qy	721	KEVYKKKEKOESEETMOQATWVYKYPYVAKHOYWKOGEEYRVYGVGMSWISKTHYRF	780
Dp	721	KEVYKKKEKOESEETMOQATWVYKYPYVAKHOYWKOGEEYRVYGVGMSWISKTHYRF	780
Qy	781	VPKLPGNTNMYRKSLEGTNNMNDENMDESDKRCGRSPKTIKIPDSKDXBKVSADAK	840
Dp	781	VPKLPGNTNMYRKSLEGTNNMNDENMDESDKRCGRSPKTIKIPDSKDXBKVSADAK	840
Qy	841	GADONENDISKITEKQOVDKELLDSDSPCKKEEEMEVODDMKTESHNCQESSQOVV	900
Dp	841	GADONENDISKITEKQOVDKELLDSDSPCKKEEEMEVODDMKTESHNCQESSQOVV	900
Qy	901	NVSEGFHLRTSYKKKTKSSKLDGLBERRIKQFTLEEKORLEKIKLEGGIKIGIKTSTNS	960
Dp	901	NVSEGFHLRTSYKKKTKSSKLDGLBERRIKQFTLEEKORLEKIKLEGGIKIGIKTSTNS	960
Qy	961	KNLSESPVITRAKGCOSDMROBOSPNANNOPEDLLOGCOSOSBSVYLAMDPSHTNN	1020
Dp	961	KNLSESPVITRAKGCOSDMROBOSPNANNOPEDLLOGCOSOSBSVYLAMDPSHTNN	1020
Qy	1021	LYPDRVLDVDSISPECTKPKONSTIENDIEEVSDLASRGOEPTKSKTKGNDFTIDSK	1080
Dp	1021	LYPDRVLDVDSISPECTKPKONSTIENDIEEVSDLASRGOEPTKSKTKGNDFTIDSK	1080
Qy	1081	LASADDIGTLCKNKKRLLIOEESDTIVSSSKALSHSVKSTNDRDAPPLSRAMDFEKL	1140
Dp	1081	LASADDIGTLCKNKKRLLIOEESDTIVSSSKALSHSVKSTNDRDAPPLSRAMDFEKL	1140
Qy	1141	GCDSBSNSTLENSSDIVSIDSSEEDMIVONSNEISIQOFTREQDVEVLEPLKELCVSG	1200

Db	1141	GCDSESNSTJLENSSDIVSIJODSSEEMIVQNSNESTSEDFRREDDEVLEPLKCELVSQ	1200
Qy	1201	ESTONCEDRLPVKGTBEANGKRPSPQOKLEERPYNKCSDOJIKLNTDCKNNENRESEKG	1260
Db	1201	ESTONCEDRLPVKGTBEANGKRPSPQOKLEERPYNKCSDOJIKLNTDCKNNENRESEKG	1260
Qy	1261	QRTSTPDJNKKDKPKIYILKGECLKRISSRVSGNVEPKYNNLNKIIPENDIKSLVKE	1320
Db	1261	QRTSTPDJNKKDKPKIYILKGECLKRISSRVSGNVEPKYNNLNKIIPENDIKSLVKE	1320
Qy	1321	SAIRPFINGVIMEDENERNRSSETKSHLLSSSDAEGNRRDSELTLPYKRESDSTQTTTPS	1380
Db	1321	SAIRPFINGVIMEDENERNRSSETKSHLLSSSDAEGNRRDSELTLPYKRESDSTQTTTPS	1380
Qy	1381	ASCPESNSVNOVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNENVGES	1440
Db	1381	ASCPESNSVNOVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNENVGES	1440
Qy	1441	KRKRVITEVYTMSTYAETSEKYIKERKDKQTVVSSSTENACKSVTTTTTTPVKLSTPS	1500
Db	1441	KRKRVITEVYTMSTYAETSEKYIKERKDKQTVVSSSTENACKSVTTTTTTPVKLSTPS	1500
Qy	1501	TGGSVDJIIISVKEOSKPVTTTTVDSLTITGGJLYVMSMTSEKSEYTRDKVKLTKMKSRRPKT	1560
Db	1501	TGGSVDJIIISVKEOSKPVTTTTVDSLTITGGJLYVMSMTSEKSEYTRDKVKLTKMKSRRPKT	1560
Qy	1561	RSGTALPSYRKFPYTKSTKKSIFVLPNDDIKLKLARKGGLREVPYFNYNKAPALDIWPIRSP	1620
Db	1561	RSGTALPSYRKFPYTKSTKKSIFVLPNDDIKLKLARKGGLREVPYFNYNKAPALDIWPIRSP	1620
Qy	1621	RPTGCTIMWRRLQTVVSLAGVSIJMLLEMLAASIRBMDMAKXVPPGGGSRTESETETTT	1680
Db	1621	RPTGCTIMWRRLQTVVSLAGVSIJMLLEMLAASIRBMDMAKXVPPGGGSRTESETETTT	1680
Qy	1681	EIKRRDVGXYGLRFEYCJIKRICKPIGVEPEKRETPTPQRKLRBSALRPRKREPQKTG	1740
Db	1681	EIKRRDVGXYGLRFEYCJIKRICKPIGVEPEKRETPTPQRKLRBSALRPRKREPQKTG	1740
Qy	1741	PVLIETWVAEELELMEIRAFABRVEKEKAQAVEQAKKREJQOKPVIAIATSTTSPST	1800
Db	1741	PVLIETWVAEELELMEIRAFABRVEKEKAQAVEQAKKREJQOKPVIAIATSTTSPST	1800
Qy	1801	TSTISPRQAKWVAPISSSVTGTCKMVLTRKVSAPVTFPOQKNKHOFATWVWOGOSNS	1860
Db	1801	TSTISPRQAKWVAPISSSVTGTCKMVLTRKVSAPVTFPOQKNKHOFATWVWOGOSNS	1860
Qy	1861	GVVVOOVOKVIGIIPSSSTGSOQTFSTFOPRATVITIRPMTSSGGTTSNSOVIITGPQIRP	1920
Db	1861	GVVVOOVOKVIGIIPSSSTGSOQTFSTFOPRATVITIRPMTSSGGTTSNSOVIITGPQIRP	1920
Qy	1921	GMVIVIRPLQOSTLGKAIIRTPVWVOPGAPQOVMTQIIRGQVSTAVSAPMTVSSTPGOK	1980
Db	1921	GMVIVIRPLQOSTLGKAIIRTPVWVOPGAPQOVMTQIIRGQVSTAVSAPMTVSSTPGOK	1980
Qy	1981	SLTSTSTSNIOSSASOPRPRPOGOVKYKLMALQJLTQJSHGNSOGLJYVIOGQGTJQOL	2040
Db	1981	SLTSTSTSNIOSSASOPRPRPOGOVKYKLMALQJLTQJSHGNSOGLJYVIOGQGTJQOL	2040
Qy	2041	QJIRPOGTVILPBGPOQJLMOQAMPNGVORLEPLPLATTTTASTTTTTVSTTAAGTGBOR	2100
Db	2041	QJIRPOGTVILPBGPOQJLMOQAMPNGVORLEPLPLATTTTASTTTTTVSTTAAGTGBOR	2100
Qy	2101	QSKSLSPOMQONODKTLPRPAOSSSVSGAKKAPQOTAPRSARPCQOTOPSPAPEVOTQREY	2160
Db	2101	QSKSLSPOMQONODKTLPRPAOSSSVSGAKKAPQOTAPRSARPCQOTOPSPAPEVOTQREY	2160
Qy	2161	QJOTQTVSSHVSEAPQHTHAOSSKPVYAAOSSQPSNVOGOSPVYVOSPSQTRIRPSTPSQL	2220
Db	2161	QJOTQTVSSHVSEAPQHTHAOSSKPVYAAOSSQPSNVOGOSPVYVOSPSQTRIRPSTPSQL	2220
Qy	2221	SPGQOSOVQTTTSPJPIQHTSLQJLPSOGQOQSQPOVQVQSSQJTLSSQJLNOYVSVSPS	2280
Db	2221	SPGQOSOVQTTTSPJPIQHTSLQJLPSOGQOQSQPOVQVQSSQJTLSSQJLNOYVSVSPS	2280

Db 2221 SGGQSOVOTTTSSQPIPIOPHTSLQIPSGQPOSDPOVOSTQTUJLSSGOTLNOVSYSFS 2280
Qy 2281 RPOLOIOOPQOVIAVPPOLQOQOVVLSQIQSOVVAQIQAOSSGVPQOIKLOLPIQIQOSS 2340
Db 2281 RQQLDIOQOPQOVIAVPPOLQOQOVVLSQIQSOVVAQIQAOSSGVPQOIKLOLPIQIQOSS 2340
Qy 2341 AVQTHQIQNVTVQAASVOEQLOQVQQLDQOQKQOQOIEIKREHTLOASNOSEITIQO 2400
Db 2341 AVQTHQIQNVTVQAASVOEQLOQVQQLDQOQKQOQOIEIKREHTLOASNOSEITIQO 2400
Qy 2401 VVMKHNVAVEHLKOKKSMTPAREENQBMIVONWKKYITLIDIKREKKAARKKREESV 2460
Db 2401 VVMKHNVAVEHLKOKKSMTPAREENQBMIVONWKKYITLIDIKREKKAARKKREESV 2460
Qy 2461 EOKRSKQNTKLSALLFKHEKEDLRAEILKRALDLKDLOIEVOEELKRLKIKKEKDLQ 2520
Db 2461 EOKRSKQNTKLSALLFKHEKEDLRAEILKRALDLKDLOIEVOEELKRLKIKKEKDLQ 2520
Qy 2521 LAQATAVAAPCPPVTPVLPAPAPPPSPPPPGVQHTGLSTPTLPVASOKRREBEKDS 2580
Db 2521 LAQATAVAAPCPPVTPVLPAPAPPPSPPPPGVQHTGLSTPTLPVASOKRREBEKDS 2580
Qy 2581 SSKSKKKMISTTSKETKDTLYCICKTPYDESKFYICGDCRQNNYHRCVGILOSEAE 2640
Db 2581 SSKSKKKMISTTSKETKDTLYCICKTPYDESKFYICGDCRQNNYHRCVGILOSEAE 2640
Qy 2641 LIDEVYCPOCOSTEDAMVLTPLTEKDEYGLKRVLSLOAHKMAPFLPVPDNDADY 2700
Db 2641 LIDEVYCPOCOSTEDAMVLTPLTEKDEYGLKRVLSLOAHKMAPFLPVPDNDADY 2700
Qy 2701 GVIKERMPLATMEERVORRYEKLTEFVADMTKIEDNCRYNPSDPFYQCAEVLESFV 2760
Db 2701 GVIKERMPLATMEERVORRYEKLTEFVADMTKIEDNCRYNPSDPFYQCAEVLESFV 2760
Qy 2761 OKLGFKASRSHNNKLQSTAS 2781
Db 2761 OKLGFKASRSHNNKLQSTAS 2781

RESULT 2
US-09-698-295-1
: Sequence 1, Application US/09698295
: GENERAL INFORMATION:
: APPLICANT: Jones, Michael H.
: TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
: FILE REFERENCE: 06501-068001
: CURRENT APPLICATION NUMBER: US/09/698,295
: CURRENT FILING DATE: 2000-10-27
: PRIOR APPLICATION NUMBER: PCT/JP99/02340
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: JAPAN 10/137631
: PRIOR FILING DATE: 1998-04-30
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2907
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-698-295-1

Query Match 98.9%; Score 14170; DB 20; Length 2907;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;

Qy 1 MYSEEEEDDGAETQOSEDEDEDEDEDDDDSDYPREMEDDDDDASCTESSFSHST 60
Db 1 MYSEEEEDDGAETQOSEDEDEDEDEDDDDSDYPREMEDDDDDASCTESSFSHST 60
Qy 61 YSSTGRKKPRVHRRSPILKEKDIPLLEFPKSSBDLAMPNEHIMNVIAIYELRNFGTV 120
Db 61 YSSTGRKKPRVHRRSPILKEKDIPLLEFPKSSBDLAMPNEHIMNVIAIYELRNFGTV 120
Qy 121 LRLSPREFDFCALVSOEQTLMAEMHVVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180

Db 121 LRLSPREFDFCALVSOEQTLMAEMHVVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
Qy 181 FIDGMWPEVLRVYCSDSDEYHNVLPYOEAEQYVGVPEVNEKIKVLQELVDQFLTITIAE 240
Db 181 FIDGMWPEVLRVYCSDSDEYHNVLPYOEAEQYVGVPEVNEKIKVLQELVDQFLTITIAE 240
Qy 241 ELMSEGVIOYDHCRCVCHLGLDLCCETCSAVYHLEBCVAPRLEEVEDEMQCEVCANHY 300
Db 241 ELMSEGVIOYDHCRCVCHLGLDLCCETCSAVYHLEBCVAPRLEEVEDEMQCEVCANHY 300
Qy 301 PGVTDCAVEIOKNKPYIRHEPIGYDRSRKRYWFLNRLLIIEEDTENENEKIMVYSTKYQ 360
Db 301 PGVTDCAVEIOKNKPYIRHEPIGYDRSRKRYWFLNRLLIIEEDTENENEKIMVYSTKYQ 360
Qy 361 LVELDLCQDWEALCQILPEMREIHRHMDITEDLTKNKGSKSFLAANEETLS 420
Db 361 LVELDLCQDWEALCQILPEMREIHRHMDITEDLTKNKGSKSFLAANEETLS 420
Qy 421 IRAKKGIDINVASPEETEKDKNETENDSKDAEKNREEFEDQSLKSDDKTPDDDEQOK 480
Db 421 IRAKKGIDINVASPEETEKDKNETENDSKDAEKNREEFEDQSLKSDDKTPDDDEQOK 480
Qy 481 SE----- 482
Db 481 SEPTREVGDKGNSVANLGDNTTNNATSEETSPSEGRSPVGLSETPDSSNMAEKVASEL 540
Qy 483 ----- 482
Db 541 PODVPEPNKTCSSNTSATTTSIQPLENSSSELSNQSSESAAADDPENGERSHT 600
Qy 483 -----VGFSEKSEKSNELSESPPAGKASGSTRILIRLRNPDSKLSQKSOQVAAA 534
Db 601 PVSIOEIVGDTSEKSTGELSESPPAGKASGSTRILIRLRNPDSKLSQKSOQVAAA 660
Qy 535 HEANKLFKSGEYLVVNSOGEISRLSTKKVIMKGINNYFKLQGEKRYVYHNOYSTNS 594
Db 661 HEANKLFKSGEYLVVNSOGEISRLSTKKVIMKGINNYFKLQGEKRYVYHNOYSTNS 720
Qy 595 FALNKHQREDDHKRRHLANKPCLTPAGEFKMGSVHSGSVLITSLRLITJOLENNIS 654
Db 721 FALNKHQREDDHKRRHLANKPCLTPAGEFKMGSVHSGSVLITSLRLITJOLENNIS 780
Qy 655 SFLLPMMASHBRANWIKAYOMCSKREFALALILECAVAPVVALPIMREFLGHTRLRHMT 714
Db 781 SFLLPMMASHBRANWIKAYOMCSKREFALALILECAVAPVVALPIMREFLGHTRLRHMT 840
Qy 715 STEREEKYVKKKKKQOEEETWQATWVKYTFPVKHQVWKQGEYRVYGYGQMSWISK 774
Db 841 STEREEKYVKKKKKQOEEETWQATWVKYTFPVKHQVWKQGEYRVYGYGQMSWISK 900
Qy 775 THYVRFVPRKLPQNTNANNYKSLBGTNNMNDENDESDKRCSPKKIKIIEPSEKDEVK 834
Db 901 THYVRFVPRKLPQNTNANNYKSLBGTNNMNDENDESDKRCSPKKIKIIEPSEKDEVK 960
Qy 835 GSDAAGADQNEMDISKITEKKQDVYKELLDSDSKPCKEPEAEVDDDKTESHVACQES 894
Db 961 GSDAAGADQNEMDISKITEKKQDVYKELLDSDSKPCKEPEAEVDDDKTESHVACQES 1020
Qy 895 SOYDVVNVSEGFHLRTSYKKTSSKLDGLLERRIKQFTLEEKQRIEIKLEGGIKIGK 954
Db 1021 SOYDVVNVSEGFHLRTSYKKTSSKLDGLLERRIKQFTLEEKQRIEIKLEGGIKIGK 1080
Qy 955 TSTNSSKNLSSESPVITKAKEGQSDSMRQDSNNANDQPEDLIQCSQSDSSVLRMSDP 1014
Db 1081 TSTNSSKNLSSESPVITKAKEGQSDSMRQDSNNANDQPEDLIQCSQSDSSVLRMSDP 1140
Qy 1015 SHTTNKLVPKRVLDVYSIRSPETKCPKQNSIENDIEEYVSDLASGQPTSKTKGNF 1074
Db 1141 SHTTNKLVPKRVLDVYSIRSPETKCPKQNSIENDIEEYVSDLASGQPTSKTKGNF 1200
Qy 1075 FTIDSKLASADDIGTLICKNKKPLIQEESDTIVSSKSALHSSVPKSTNDRAATPLSRAM 1134


```

Db 1201 FIDSKLASADIGTLICNNKKPLIOESDPTIVSSSKSALHSSVSPKSTNDROATPLSRAM 1260
QY 1135 DEFGKJGCSSEMSSTLNSDPTVSIODSSSEDMITVONSNSEISTEORTREROVELEPLK 1194
Db 1261 DEFGKJGCSSEMSSTLNSDPTVSIODSSSEDMITVONSNSEISTEORTREROVELEPLK 1320
QY 1195 CELVSESTGNCBDRPLPVKQTEFANGKRPQSOQKLEERPVKCSDOJIKLNTTKNNENR 1254
Db 1321 CELVSESTGNCBDRPLPVKQTEFANGKRPQSOQKLEERPVKCSDOJIKLNTTKNNENR 1380
QY 1255 ESEKQKORTSTFOINGKDNKPKIYLKGECLKEISESRVVSNGNVEPYNNINKIIPENDIK 1314
Db 1381 ESEKQKORTSTFOINGKDNKPKIYLKGECLKEISESRVVSNGNVEPYNNINKIIPENDIK 1440
QY 1315 SLVKSASAIKPFINGDIYMEDEFERNSSSETKSHLSSSDAEGNRPDLSTPSTKESDPT 1374
Db 1441 SLVKSASAIKPFINGDIYMEDEFERNSSSETKSHLSSSDAEGNRPDLSTPSTKESDPT 1500
QY 1375 QTTTPSACPESSNVNOVEDMEIETSEVKKVTSSPTISEESNLNDFTIDENGLPTINKNE 1434
Db 1501 QTTTPSACPESSNVNOVEDMEIETSEVKKVTSSPTISEESNLNDFTIDENGLPTINKNE 1560
QY 1435 NVNGESKRTVITEVYTMSTVATVETSEKTYIKVEKGDQOTVVSSTENCAKSTVTTTTVT 1494
Db 1561 NVNGESKRTVITEVYTMSTVATVETSEKTYIKVEKGDQOTVVSSTENCAKSTVTTTTVT 1620
QY 1495 KLSPTSGSVDIISYKESKTYVTVTDSLTFTTGGLVTSMTVSEKSTYTRDKVKLMKF 1554
Db 1621 KLSPTSGSVDIISYKESKTYVTVTDSLTFTTGGLVTSMTVSEKSTYTRDKVKLMKF 1680
QY 1555 SRPKRTSGTALPSYKREKVTSTKRSIFVLPNDDLKKLARKGGIREVYFNNAKPALDI 1614
Db 1681 SRPKRTSGTALPSYKREKVTSTKRSIFVLPNDDLKKLARKGGIREVYFNNAKPALDI 1740
QY 1615 WPPSPRPSTGCTWRYALQVVKSLAGVSLMLRLMLASLRDDMAAKVPCGGSTRTPTSE 1674
Db 1741 WPPSPRPSTGCTWRYALQVVKSLAGVSLMLRLMLASLRDDMAAKVPCGGSTRTPTSE 1800
QY 1675 TEITTTTIIKRDVGPXGIFEXYCIKRIICPIGVETPKETPTPORKGLSSALRPKPE 1734
Db 1801 TEITTTTIIKRDVGPXGIFEXYCIKRIICPIGVETPKETPTPORKGLSSALRPKPE 1860
QY 1735 TPKQTGVIIEYVVAEELELMEIRAFAEVREKEKAQAVEQAKKRLBOQKPVVIATSTT 1794
Db 1861 TPKQTGVIIEYVVAEELELMEIRAFAEVREKEKAQAVEQAKKRLBOQKPVVIATSTT 1920
QY 1795 SPTSSTSTSPAKVWAVAPISGSVTTGTVMVLTTKVGSFATVTFQONKNFHOTFAWVK 1854
Db 1921 SPTSSTSTSPAKVWAVAPISGSVTTGTVMVLTTKVGSFATVTFQONKNFHOTFAWVK 1980
QY 1855 QGOSNSGVVOQKVLGIISSSTGTSQOTFTSFQPRATVYIRPNTSGSGGTTSNSOVIT 1914
Db 1981 QGOSNSGVVOQKVLGIISSSTGTSQOTFTSFQPRATVYIRPNTSGSGGTTSNSOVIT 2040
QY 1915 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVAVQCAPQOVMTQIIRGQPVSTAVSAPNTVS 1974
Db 2041 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVAVQCAPQOVMTQIIRGQPVSTAVSAPNTVS 2100
QY 1975 STPOKSLTANTSTNSIOSSASQPPRPQOGOVKLTMAQLTQLTQGHGNGCLTVVIOGOG 2034
Db 2101 STPOKSLTANTSTNSIOSSASQPPRPQOGOVKLTMAQLTQLTQGHGNGCLTVVIOGOG 2160
QY 2035 QTTGOLILPOGVTVLPGPQOQLQOAMAPNGTVORFLFTPLATTATATSTTTTSTTAA 2094
Db 2161 QTTGOLILPOGVTVLPGPQOQLQOAMAPNGTVORFLFTPLATTATATSTTTTSTTAA 2220
QY 2095 GTGBOROSKLSPOMVHODKTLPPAQSSSVGPAKAQOPQAPQAPQOTQPOSPAPQEV 2154
Db 2221 GTGBOROSKLSPOMVHODKTLPPAQSSSVGPAKAQOPQAPQAPQOTQPOSPAPQEV 2280
QY 2155 QTOPEVOTQTTVSSHVSEAKOPHTHAOSSKPOVAAQSQPOSNVQOQSFRVRSQOTRIRP 2214
Db 2281 QTOPEVOTQTTVSSHVSEAKOPHTHAOSSKPOVAAQSQPOSNVQOQSFRVRSQOTRIRP 2340

```

```

QY 2215 STPSQLSPGOQSOVQTTTSQPIPIQPHTSLOIPSGQPOQPOQSOPOVOSSTQTLSSGQTLNOV 2274
Db 2341 STPSQLSPGOQSOVQTTTSQPIPIQPHTSLOIPSGQPOQPOQSOPOVOSSTQTLSSGQTLNOV 2400
QY 2275 SVSSPSRPOLOIQPOPOVYIAPVPOLOQOYVUSIOTOSVYVAQIOAOQSGVPQOIKLOLPI 2334
Db 2401 SVSSPSRPOLOIQPOPOVYIAPVPOLOQOYVUSIOTOSVYVAQIOAOQSGVPQOIKLOLPI 2460
QY 2335 QIOOSSAVOHTOIONVYVTOAASVBOLORVQOLRPOQOKKKQOQOIEIKREHTLOASNOS 2394
Db 2461 QIOOSSAVOHTOIONVYVTOAASVBOLORVQOLRPOQOKKKQOQOIEIKREHTLOASNOS 2520
QY 2395 EIIOQOVYKHNVAIEHLKOKKSMTPAEREENQRMIVCQVMKYILDKIDKEEQAKKR 2454
Db 2521 EIIOQOVYKHNVAIEHLKOKKSMTPAEREENQRMIVCQVMKYILDKIDKEEQAKKR 2580
QY 2455 KREESVQKRSKONATKLSALLEKHKQURAEILKRLALLDKLOLQIEVQDELRDLKIKK 2514
Db 2581 KREESVQKRSKONATKLSALLEKHKQURAEILKRLALLDKLOLQIEVQDELRDLKIKK 2640
QY 2515 EKDLMOLOATVAAPACPPVTPVLPAPAPPPSPPPPGVQHTGLSTPTLPVYASOKRRR 2574
Db 2641 EKDLMOLOATVAAPACPPVTPVLPAPAPPPSPPPPGVQHTGLSTPTLPVYASOKRRR 2700
QY 2575 EEEKDSSSKSKKKKMISTTSKETKKDKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI 2634
Db 2701 EEEKDSSSKSKKKKMISTTSKETKKDKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI 2760
QY 2635 LOSEAEILDEYVCPQOSTDANTVLTPLTEKYEGLKRVLRSLQAHKMAWPLEFVDN 2694
Db 2761 LOSEAEILDEYVCPQOSTDANTVLTPLTEKYEGLKRVLRSLQAHKMAWPLEFVDN 2820
QY 2695 DADYGVYIKEMDLATMEERVORRYEKLFEVADMTKIFDCRCRYVNSDSFPYOCABV 2754
Db 2821 DADYGVYIKEMDLATMEERVORRYEKLFEVADMTKIFDCRCRYVNSDSFPYOCABV 2880
QY 2755 LESFVQKLGFKRASRSHNNKLOSTAS 2781
Db 2881 LESFVQKLGFKRASRSHNNKLOSTAS 2907

RESULT 3
PCT-US02-21179-15
; Sequence 15, Application PC/TUS0221179
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: CANDHI, Ameena R.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YUE, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: GRIPPIN, Jennifer A.
; APPLICANT: YANG, Junming
; APPLICANT: SANJANMALA, Madhu M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BOROMSKY, Mark L.
; APPLICANT: YAO, Monique G.
; APPLICANT: WALIA, Nalinder K.
; APPLICANT: BANDMAN, Olga
; APPLICANT: IAL, PREETI G.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LIO, Wen

```

```
APPLICANT: TANG, Y. Tom
APPLICANT: ZEBARJADIAN, Yeganeh
APPLICANT: LU, Yan
TITLE OF INVENTION: Nucleic Acid-Associated Proteins
FILE REFERENCE: PF-1031 PCT
CURRENT APPLICATION NUMBER: PCT/US02/21179
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 60/300,518
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/301,787
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,792
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,892
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,893
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/303,405
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/303,442
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/364,438
PRIOR FILING DATE: 2002-03-12
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PERL Program
SEQ ID NO 15
LENGTH: 2759
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 394992CD1
PCT-US02-21179-15

Query Match      93.7%; Score 13428; DB 1; Length 2759;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 2625; Conservative 6; Mismatches 7; Indels 264; Gaps 3;

QY 1 MYSEEEEDDAEETQDEDEDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFSHST 60
DB 1 MYSEEEEDDAEETQDEDEDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFSHST 60
QY 61 YSSTPGRKRRPRVHRPRSP1LEEKDIRPLEPPKSSSEDLMPVNEHINVAIAYEVLNFGTV 120
DB 61 YSSTPGRKRRPRVHRPRSP1LEEKDIRPLEPPKSSSEDLMPVNEHINVAIAYEVLNFGTV 120
QY 121 LRLSPREFDFCAALYSOQCTLMAEMHYLLKAVLREEDTSTTGPADLKQSVNSTLY 180
DB 121 LRLSPREFDFCAALYSOQCTLMAEMHYLLKAVLREEDTSTTGPADLKQSVNSTLY 180
QY 181 FTDGMTWPEVLYRYCSBDEKEYNHVLRYQEAEDYRYGVENKIKVLOFLVDQFLTTNIAEE 240
DB 181 FTDGMTWPEVLYRYCSBDEKEYNHVLRYQEAEDYRYGVENKIKVLOFLVDQFLTTNIAEE 240
QY 241 ELMSBSGVIOYDHCRCVCHKLGLDLCCETCSAYUHEBCVYKPLEEYVDEDMQCEVYAHKY 300
DB 241 ELMSBSGVIOYDHCRCVCHKLGLDLCCETCSAYUHEBCVYKPLEEYVDEDMQCEVYAHKY 300
QY 301 PCVTDCAVETQKNKPYIRHREPIGYDSRRKYWFNLRLIIEEDTENENKKIUYSTKYQ 360
DB 301 PCVTDCAVETQKNKPYIRHREPIGYDSRRKYWFNLRLIIEEDTENENKKIUYSTKYQ 360
QY 361 LAELIDCLDKDYWEALCKILEEMREIRHMDITEDLTNKAKGSKSFLLAANEELLES 420
DB 361 LAELIDCLDKDYWEALCKILEEMREIRHMDITEDLTNKAKGSKSFLLAANEELLES 420
QY 421 IIAKGGIDINVKSPETEKKDKNETENDSKDAEKNREEFEDQSLKSDSDKTPDDPEQCK 480
DB 421 IIAKGGIDINVKSPETEKKDKNETENDSKDAEKNREEFEDQSLKSDSDKTPDDPEQCK 480
QY 481 SE----- 482
DB 481 SEEPTEVGDKGNVSANLQDNFTTNAETSEETSPSEGRSPVGCLETDPDSSNMAKKVASEL 540
```

```
QY 483 -----VDFEKSXNGELSESPGAGSGSTRITRLRNPDLSQLSQVAAAAHEA 537
DB 541 PDVPYGDGFSEKXNGELSESPGAGSGSTRITRLRNPDLSQLSQVAAAAHEA 600
QY 538 NKLFEKGKVELVYNSQGEISRLSTKKEVIMKGINNYFKLQGGKRRVYHNOYSTSPAL 597
DB 601 NKLFEKGKVELVYNSQGEISRLSTKKEVIMKGINNYFKLQGGKRRVYHNOYSTSPAL 660
QY 598 NKQHEDEHDKRRHLAHKFLCPAGEFKMNGSVHGSKVLTISRLTLTQLENNISSFL 657
DB 661 NKQHEDEHDKRRHLAHKFLCPAGEFKMNGSVHGSKVLTISRLTLTQLENNISSFL 720
QY 658 HPNMAHRAHWIKAYQCMSPREFALALILECAVRYVWLPYIRREFLGHTRLRMTSIE 717
DB 721 HPNMAHRAHWIKAYQCMSPREFALALILECAVRYVWLPYIRREFLGHTRLRMTSIE 780
QY 718 REBEKRYKKKEKKQEEBETMOATWYKYTFPVKHQYWKQGEERYVYTGGSWISKTHY 777
DB 781 REBEKRYKKKEKKQEEBETMOATWYKYTFPVKHQYWKQGEERYVYTGGSWISKTHY 840
QY 778 YRFVPLPGNTNRYNRSLEGTKNMDENMDESDFKRCSPKIKIIEPDSEKDEYKGSQ 837
DB 841 YRFVPLPGNTNRYNRSLEGTKNMDENMDESDFKRCSPKIKIIEPDSEKDEYKGSQ 900
QY 838 AAKGADQNEMDISKITEKKQDYKELLDSDDKPCKEEMAEVDDMKTKTSHVNCQESSQY 897
DB 901 AAKGADQNEMDISKITEKKQDYKELLDSDDKPCKEEMAEVDDMKTKTSHVNCQESSQY 960
QY 898 DVNVSEGFHLRTSYKKTKKSKLDGLLEBKIKOFTLEEKQREKTKLBGIGIKGTST 957
DB 961 DVNVSEGFHLRTSYKKTKKSKLDGLLEBKIKOFTLEEKQREKTKLBGIGIKGTST 1020
QY 958 NSSKNLSSESPVITKAKEGOSDSMRQOSPANNDDPEDLQGCSSDSSVYLMSPSHST 1017
DB 1021 NSSKNLSSESPVITKAKEGOSDSMRQOSPANNDDPEDLQGCSSDSSVYLMSPSHST 1080
QY 1018 TNKLYPKDRVLDDVSIIRSPETKCPKQNSIENDIEEVSDIASGQEPRTSKTKGNDFID 1077
DB 1081 TNKLYPKDRVLDDVSIIRSPETKCPKQNSIENDIEEVSDIASGQEPRTSKTKGNDFID 1140
QY 1078 DSKLASADIGITLICKNNKPLIOESDITVSSSKSLHSSVPKSTYNDRDATPLSRAMDE 1137
DB 1141 DSKLASADIGITLICKNNKPLIOESDITVSSSKSLHSSVPKSTYNDRDATPLSRAMDE 1200
QY 1138 GRLGDSSESNTLENSDPVSIODSSEEDMAYONSNESTSEQRTRQOYEVLEPKCEL 1197
DB 1201 GRLGDSSESNTLENSDPVSIODSSEEDMAYONSNESTSEQRTRQOYEVLEPKCEL 1260
QY 1198 VSGESTGNCEDRLPVKGTANGKRPQOKKLEERPVNKGSDQIKLANTTDKKNENRES 1257
DB 1261 VSGESTGNCEDRLPVKGTANGKRPQOKKLEERPVNKGSDQIKLANTTDKKNENRES 1320
QY 1258 KKGQRTSTQINGKDKPKIYLYKBECLKLEISBSRVVSGVNEPVNINIKIIPENDIKSLT 1317
DB 1321 KKGQRTSTQINGKDKPKIYLYKBECLKLEISBSRVVSGVNEPVNINIKIIPENDIKSLT 1380
QY 1318 VKESAIRPFLNGDVIMEDFERNSSFTKSHLSSDAEGVYRSLTLPSTKSDSTQTT 1377
DB 1381 VKESAIRPFLNGDVIMEDFERNSSFTKSHLSSDAEGVYRSLTLPSTKSDSTQTT 1440
QY 1378 TFSASCPESSNVQVDMLEIETSEVKKVYSSPTSEESNLSNDFIDENGLPINKNENNV 1437
DB 1441 TFSASCPESSNVQVDMLEIETSEVKKVYSSPTSEESNLSNDFIDENGLPINKNENNV 1500
QY 1438 GESKRTVTITEVTTMTSTVATESKTVIKVKGDKQTVVSTENCAKSTVYTTTYYTKLS 1497
DB 1501 GESKRTVTITEVTTMTSTVATESKTVIKVKGDKQTVVSTENCAKSTVYTTTYYTKLS 1560
QY 1498 TPSTGGSVDIISKESKIVVTVTVDSLTJTGGLTVTMTVSKESSTDKYKLMKFSRP 1557
DB 1561 TPSTGGSVDIISKESKIVVTVTVDSLTJTGGLTVTMTVSKESSTDKYKLMKFSRP 1620
```

Qy	1558	KKISGALPBYRKFFYKSKTKSIPLVLPNDLKLKLKKGIFBVEPYENNAKPAIDIMPY	1617
Db	1621	KKTSRGALPBYRKFFYKSKTKSIFLVPNDLKLKLKKGIFBVEPYENNAKPAIDIMPY	1680
Qy	1618	PSPPFGITWRYRLQJWVKSLAGYSMLRLMLASLFWDDMAKVPBGSGSTPTEISEPI	1677
Db	1661	PSPPFGITWRYRLQJWVKSLAGYSMLRLMLASLFWDDMAKAPPGCGITRTETSEPI	1740
Qy	1678	TTTEIKRRDVGPGYIGFEYCIKRIKICIGVPEPKETPTPOKGLRSALPKRPENPK	1737
Db	1741	TTTEIKRRDVGPGYIGSEYCIKRIKICIGVPEPKETPTPOKGLRSALPKRPENPK	1800
Qy	1738	QTGPVILIEYWAEELELMEIRAFABREVEKKAQAEQAKRLBEQKPYANISSTSP	1737
Db	1801	QTGPVILIEYWAEELELMEIRAFABREVEKKAQAEQAKRLBEQKPYANISSTSP	1860
Qy	1798	SSTSTISPAQKVMVADISSVTGKRMVLTTCVSGPAVTEFOOKNPFHPFATWVKQO	1857
Db	1861	SSTSTISPAQKVMVADISSVTGKRMVLTTCVSGPAVTEFOOKNPFHPFATWVKQO	1920
Qy	1858	SNSGVVOQKVLGIIPSSSTGTSQOQFTFSHPRTAVTIRPNTSGSGGTSNSQVITGPQ	1917
Db	1921	SNSGVVOQKVLGIIPSSSTGTSQOQFTFSHPRTAVTIRPNTSGSGGTSNSQVITGPQ	1980
Qy	1918	IRPGVITVIRPPLQOSTGKRIIRTPVMVOPGAPQOVMIOIRGQVSTVAVSAPNTVSTP	1977
Db	1981	IRPGVITVIRPPLQOSTGKRIIRTPVMVOPGAPQOVMIOIRGQVSTVAVSAPNTVSTP	2040
Qy	1978	GOKSLTSTSTNSIOSSASOPPRPOQGVKLTMAOLTJLTQSHGNGGJLVJIOQGGTT	2037
Db	2041	GOKSLTSTSTNSIOSSASOPPRPOQGVKLTMAOLTJLTQSHGNGGJLVJIOQGGTT	2100
Qy	2038	GOLOLIPQGVTVLPGPQOLOMAMENGTVORFLTPLATATTAATSTTTTTVSTTAAGTG	2097
Db	2101	GOLOLIPQGVTVLPGPQOLOMAMENGTVORFLTPLATATTAATSTTTTTVSTTAAGTG	2160
Qy	2098	FOROSKSLPQWQHQDKTLPPAOSVSGAKAKPOPTAOPSARPQOTOPQSPAQPEVOTQ	2157
Db	2161	FOROSKSLPQWQHQDKTLPPAOSVSGAKAKPOPTAOPSARPQOTOPQSPAQPEVOTQ	2220
Qy	2158	PEVOTQTTVSSHVSEKOPRPHAAOSSKPOVAAOSQPOSNQOGSPVRYOSPSOTIRPSTP	2217
Db	2221	PEVOTQTTVSSHVSEKOPRPHAAOSSKPOVAAOSQPOSNQOGSPVRYOSPSOTIRPSTP	2280
Qy	2218	SQSLSPGOOSQVQTTTSOPRIPDPTSLQIPISQGOPOQPOVQSSQTLSGOTLNOVSVS	2277
Db	2281	SQSLSPGOOSQVQTTTSOPRIPDPTSLQIPISQGOPOQPOVQSSQTLSGOTLNOVSVS	2319
Qy	2278	SPSRPQLQIQPQPOVYAVNPQLQOQVQVLSQIOSOVVAQIOAQSGVFPQIKLQPLQIQ	2337
Db	2320	SPSRPQLQIQPQPOVYAVNPQLQOQVQVLSQIOSOVVAQIOAQSGVFPQIKLQPLQIQ	2319
Qy	2338	QSAVQTHQIQNVTVQVAAVSQEQQLQRYOQQLRBDQQKKKQOQIEIKREHTLQASQSEI	2397
Db	2320	QSAVQTHQIQNVTVQVAAVSQEQQLQRYOQQLRBDQQKKKQOQIEIKREHTLQASQSEI	2319
Qy	2398	QKQYVAKHNAVIEHLKOKKSMTPAEKEENQRMIVCNOVKYILDKIDKEEKOAKKKRRE	2457
Db	2320	QKQYVAKHNAVIEHLKOKKSMTPAEKEENQRMIVCNOVKYILDKIDKEEKOAKKKRRE	2377
Qy	2458	ESVQOKSKQNAKTLASLFLKHEQJRAELTLKRALLDKQIEVOBELKDKLIKKEKD	2517
Db	2378	ESVQOKSKQNAKTLASLFLKHEQJRAELTLKRALLDKQIEVOBELKDKLIKKEKD	2437
Qy	2518	LMQJLAQATAVAACPPTVPLPAPRAPSPPPPGVQHTGLISPTPLPVASOKRKREE	2577
Db	2438	LMQJLAQATAVAACPPTVPLPAPRAPSPPPPGVQHTGLISPTPLPVASOKRKREE	2497
Qy	2578	KDSSSKKKKKMISTSKETKKDKTKYCIKPPYDESK-----	2615
Db	2498	KDSSSKKKKKMISTSKETKKDKTKYCIKPPYDESKFYIGDCLCTWMYHGEVGLITEK	2557
Qy	2616	-----FYIGDRCQNMWYHGRCVGILOSEA	2639

DB	Sequence	Accession	Score	Length	Indels	Gaps
Db	2558 EAKKMDVYICNDCKRAQGSSEELYCICRTPEYDESQFIIGCDRCQNNYHGGCVGIIQSEA	2617				
Qy	2640 ELIDELVCPQOCSTDAATVLTPLTEKDYEGSLKRVLSIAQHKRAMPFLFEPVPNDAPDY	2699				
Db	2618 ELIDELVCPQOCSTDAATVLTPLTEKDYEGSLKRVLSIAQHKRAMPFLFEPVPNDAPDY	2677				
Qy	2700 YGVIKPEPMDLATMEERVQRRYREKLTFFVADMTKIFDNCRCRYNPSDPFYQCAEVLESFF	2759				
Db	2678 YGVIKPEPMDLATMEERVQRRYREKLTFFVADMTKIFDNCRCRYNPSDPFYQCAEVLESFF	2737				
Qy	2760 VQKLGFKASRSHNNKLOSTAS	2781				
Db	2738 VQKLGFKASRSHNNKLOSTAS	2759				
RESULT 4						
; Sequence 16042, Application PC/TUS0114827						
; GENERAL INFORMATION:						
; APPLICANT: Hyseq, Inc						
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES						
; FILE REFERENCE: 21272-104						
; CURRENT APPLICATION NUMBER: PCT/US01/14827						
; PRIOR FILING DATE: 2001-05-16						
; PRIOR APPLICATION NUMBER: 09/577,408						
; NUMBER OF SEQ ID NOS: 16102						
; SOFTWARE: Custom						
; SEQ ID NO 16042						
; LENGTH: 845						
; TYPE: PRT						
; ORGANISM: Homo sapiens						
; FEATURE:						
; NAME/KEY: DOMAIN						
; LOCATION: (301)..(316)						
; OTHER INFORMATION: PHD-finger domain identified by eMATRIX, accession number						
; OTHER INFORMATION: PF00628, p-value=5.935e-09, raw score of 15.84						
; NAME/KEY: DOMAIN						
; LOCATION: (288)..(333)						
; OTHER INFORMATION: PHD-finger domain identified by Pfam, accession name PHD, E-						
; OTHER INFORMATION: value=4.9e-15, Pfam score of 63.4						
PCT-US01-14827-16042						
Query Match 23.5%; Score 3363; DB 1; Length 845;						
Best Local Similarity 82.6%; Pred. No. 6,6e-189;						
Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2;						
Qy	1 MYSEEEEDDCAEBETQDSDEDEDEDEDDDDSDYFEMEEDDDASCTESSFRSHST	60				
Db	36 MYSEEEEDDCAEBETQDSDEDEDEDEDDDDSDYFEMEEDDDASCTESSFRSHST	95				
Qy	61 YSSPTGRRKPRVHNRPSRIIEKDIPLPEPKSSSEDLAVNENHMANVIAIYELRNPGTV	120				
Db	96 YSSPTGRRKPRVHNRPSRIIEKDIPLPEPKSSSEDLAVNENHMANVIAIYELRNPGTV	155				
Qy	121 LRLSPFREDECAALVSEQCTLMAEMHVVYLKAVLREEDTSNTTGGADLKDSVSTLY	180				
Db	156 LRLSPFREDECAALVSEQCTLMAEMHVVYLKAVLREEDTSNTTGGADLKDSVSTLY	215				
Qy	181 FIDGNTWPEVLRVYCESDKEYHNVLPLYQEAEDYRYGVEENKIKVLOFLVDQFLTINARE	240				
Db	216 FIDGNTWPEVLRVYCESDKEYHNVLPLYQEAEDYRYGVEENKIKVLOFLVDQFLTINARE	275				
Qy	241 ELMSSGVLYQYDDHRCVCHKLGDLLCCETGSVYVHLCEVKKPPLFEPVDEMOCEVCVAHKY	300				
Db	276 ELMSSGVLYQYDDHRCVCHKLGDLLCCETGSVYVHLCEVKKPPLFEPVDEMOCEVCVAHKY	335				
Qy	301 PGVTDCAVEIQRNRPYRHPYIGYDRSRKRYWFLNRRLLIEEDTENDNEKKIYYSTKVO	360				
Db	336 PGVTDCAVEIQRNRPYRHPYIGYDRSRKRYWFLNRRLLIEEDTENDNEKKIYYSTKVO	395				
Qy	361 LAELIDLCADKQYWEAELOCKLIEEMRREIHRHMDITTEDLTFKAGSGNSKFLAANEEILES	420				

```
Db 396 LAELDLCDLCKDWEALCKITLSEMRKEJHRMDITEDLTNRKAGSKVSLAANEETLES 455
Oy 421 IRAKGGIDNVKSPETEKDKNETENDSKDAEKNREEFEDQSLKSDDKTPDDPEQCK 480
Db 456 IRAKGGIDNVKSPETEKDKNETENDSKDAEKNREEFEDQSLKSDDKTPDDPEQCK 515
Oy 481 SE----- 482
Db 516 SEPTVEVGDKNVSANLGDNTNATSEETSPSEGRPVYGCLETPDSSNMAEKVASEL 575
Oy 483 ----- 482
Db 576 PODVEEPNKTCESNTSATTTTSTIOPNLNSNSSSELSSQESAKAADDPENGERSHT 635
Oy 483 -----VGFSEKSEKNGELSESPGAGKAGSSTRITRLNPNPSKLSQOYAAA 534
Db 636 PYSIOEIVGDTSEKSTGELSESPGAGKAGSSTRITRLNPNPSKLSQOYAAA 695
Oy 535 HEANKLFEGKEVLVNSOGELISRLSTKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS 594
Db 636 HEANKLFEGKEVLVNSOGELISRLSTKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS 755
Oy 595 FALNKHQREDHDKRRHLAKKCLTPPAGEFKNGSVHGSKVLTITSLRITTOLENNIPS 654
Db 756 FALNKHQREDHDKRRHLAKKCLTPPAGEFKNGSVHGSKVLTITSLRITTOLENNIPS 811
Oy 655 SFLHPNM 661
Db 812 TSLHPSF 818
```

RESULT 5

US-09-629-469A-14273

; Sequence 14273, Application us/09629469A

; GENERAL INFORMATION:

```
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629, 469A
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159, 590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183, 322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14273
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-629-469A-14273
```

Query Match 18.8%; Score 2695.5; DB 20; Length 557;
Best Local Similarity 79.0%; Pred. No. 8,4e-150;
Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;

```
Oy 1811 MWAPISGVYTGCTKVLTKVGSPPATVTFQOKNKHQEPATVWKGQSSGVVVOQKVL 1870
Db 1 MWAPISGVYTGCTKVLTKVGSPPATVTFQOKNKHQEPATVWKGQSSGVVVOQKVL 60
Oy 1871 GIIPSTGTSGQTFSTFOPRTAVTIRPNTSGSGGTTSSNQYITGPQIRPGMTVIRPLQ 1930
Db 61 GIIPSTGTSGQTFSTFOPRTAVTIRPNTSGSGGTTSSNQYITGPQIRPGMTVIRPLQ 120
Oy 1931 OSTLGKAILIRTPVWQPGAPQGVMTQIIRGQPVSTAVSAPNTVSSTPGOKSLTSATSTSN 1990
Db 121 OSTLGKAILIRTPVWQPGAPQGVMTQIIRGQPVSTAVSAPNTVSSTPGOKSLTSATSN 180
Oy 1991 IOSSASQPPRPOGGQVKLTMAQLTOLTOGHGNGOGLTVYIOGGQTTGLOLIRPGVYVL 2050
Db 181 IOSSASQPPRPOGGQVKLTMAQLTOLTOGHGNGOGLTVYIOGGQTTGLOLIRPGVYVL 240
Oy 2051 PGPQGOQLMAAMPNGTVORFLFTPLATTATTTTSTTAAGTGEOROSKLSPOKOV 2110
Db 241 PGPQGOQLMAAMPNGTVORFLFTPLATTATTTTSTTAAGTGEOROSKLSPOKOV 300
Oy 2111 HODKTLPPAQSSTSVGPAPAKAPQTAQPSARPOPTQPSAPQEVOTQPEVOTQTVSSHV 2170
Db 301 HODKTLPPAQSSTSVGPAPAKAPQTAQPSARPOPTQPSAPQEVOTQPEVOTQTVSSHV 360
Oy 2171 PSEAOPTHAQSSKPVYAAQSQPSQSNVQGSQPVYVQSPSOTRIRPSTPSQLSGQSSQVOT 2230
Db 361 PSEAOPTHAQSSKPVYAAQSQPSQSNVQGSQPVYVQSPSOTRIRPSTPSQLSGQSSQVOT 420
Oy 2231 TTSOPRIPQPHSTLQIPSGQPOSQPOVQSSQTQLSSGQTLNQVSVSPSRQLQIQOPQ 2290
Db 421 TTSOPRIPQPHSTLQIPSGQPOSQ----- 446
Oy 2291 PQYIAPVQLQGVVLSQISQSVAAQIQAGSQVPOQIKLQLPIQIQSSAVQTHQIONV 2350
Db 447 ----- 446
Oy 2351 VTVQAASVQEQIQRVQQLRDQOQKKQOQIEIKREHTLOASQSEITQKOVYKKNAYIE 2410
Db 447 -----QVYKKNAYIE 457
Oy 2411 HLKOKSMTPAERENQRTVCNOYMYTILDKIDKEKOAAKKRRESVEKRSKONAT 2470
Db 458 HLKOKSMTPAERENQRTVCNOYMYTILDKIDKEKOAAKKRRESVEKRSKONAT 517
Oy 2471 KLSALLFKHKEQLRAEILKKRALLDKLDIQIEVQEEIKRDL 2510
Db 518 KLSALLFKHKEQLRAEILKKRALLDKLDIQIEVQEEIKRDL 557
```

RESULT 6

US-09-614-150-32997

; Sequence 32997, Application us/09614150

; GENERAL INFORMATION:

```
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: C1000728
; CURRENT APPLICATION NUMBER: US/09/614, 150
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157, 832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160, 191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161, 932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164, 769
; PRIOR FILING DATE: 1999-11-12
```



```
Db 517 -----GDASESOSIE-----GTRKOECKKVTN-----OKSNOLT----- 547
Oy 539 KLFKEGKEVLVYNSOGELISRLSTKKEVIMKGNINNYEKLGOEGKYRVYHNOYSTNSFALN 598
Db 548 -----NGLHFKLMEGEGFNYYNOYSTNFIALN 576
Oy 599 KQHREDHDKRRHLAHKFLTPAGEFKWNGSVHGSVLTISTRLTLITOLENNIPSSFLH 658
Db 577 KQORNEERDKRRHLSHKFLTASDFKWTGTMGTDMNITTLRQTLINFEESNIAASFILN 636
Oy 659 PNMASHRAWIKAVQCSKPREPALALALIECAVKNPVYMLPIREELGTRHLRHMSIER 718
Db 637 IMVYVKKKIMNAAVNARPSSEFAVLLLPQASLKSVEFANVWHEDLGHTLORITSAER 696
Oy 719 EEKEVKKKKKKKO---EEBETMQATWVKTTPVVKHQVWKQKEEYRVYGGGSMVSKT 775
Db 697 EERKKLEKREKREDEEERNNRLAFNYIKYTLGLKHQVWKQKEEYRVHGWMLWSS 756
Oy 776 -----HYFVFKPLPGNTVYVNRKSLGTRKNMNMENMDESCKRCSRPKKIKIEPDSEK 830
Db 757 RRCGVARRAQPLTHNRVYVHYTM---GEENDVNEIIT---LVDPRTQRFMOQCES 805
Oy 831 DEYKSDAAKAGADQ--NEMDISKITEKKDQVYKELSDSDCKPEPEVDDMKTESH 888
Db 806 SNVDGQVCHYLPDQYKNVYIEDVTE-----KIKGH 836
Oy 889 VNCQESSQYDVVNVSEGFHL--RTSYKKKTKSKLDGLERIKOTPLEKORLEKIKLE 946
Db 837 -----IDVSKALNAPGRTYYSKVAARKSRDLDDLDRRLKLAIEVEQOM---ASKIP 882
Oy 947 GGIGIGIKTSTNSSKNLSES 966
Db 883 SDMKPLVSSQNNNTANSKOT 902

RESULT 8
US-60-191-681-25768
; Sequence 25768, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: LI, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25768
; LENGTH: 976
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-681-25768
```

```
Query Match 10.9%; Score 1557.5; DB 27; Length 976;
Best Local Similarity 36.4%; Pred. No. 3,1e-82;
Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;
```

```
Oy 4 EEEEDDGAETOSDEDEDEDEDDSDOYPREMEDDDDDASCTSSSPRSHSTYSS 63
Db 99 EEEYHXSDFGSEEDKSDNEDMDLLTPSDDESLEAVNSESEFSVC---SFQKNCV--- 152
Oy 64 TPGRRKPRVHRPSPILIEE-KDIPLEFPKSSDDLAVPNEHIMNVAVIAYEVLNFGTVLR 122
Db 153 ---GRRPRRSPPEYWLQGRQYALDLDPSSSEDLFIANTHYLRALSTIEVLRFRFHMVR 209
Oy 123 LSPFRFEDCALVSOECTLAEEMHVLLKAVLREEDTSNTTFGPADIKDSVNSTLYET 182
Db 210 LSPFRFEDCALACEESGALILEVIMLKAILREDAQGTIFGFLDQKDTVINISLYLI 269
Oy 183 DGMTPEVLRVYCSKDKEY---HHVLRQYQADYDPYGVENEKKIVYLOFLVQOFLTTNLA 238
Db 270 DSITPEVLRVYCSKDKEY---HHVLRQYQADYDPYGVENEKKIVYLOFLVQOFLTTNLA 238
```

```
Oy 239 REELMSGEVLYDXDHCRVCHKIGDLLCCETCSAVYHLECYKPPLEEVPEDEMOCEYVAH 298
Db 327 RDVMLQEGPIHYHDHCRVCHRLGDLCCETCIPAVYHLECYVDPMDVYPTEDMQCGCLRSH 386
Oy 299 KVPYVDCYAEIOKKNPYIRHEPIGYDRSRKRYFPLNRLIIEDEPNEKEKIWTYSK 358
Db 387 KXGVVDVCLPOEKQGVLLRHDSLGVRHGRKRYFARLIFI-EDENFP--CWYSTT 442
Oy 359 VOLAEILDLCKDYWEAELECKTIEEMREFIHRMDTDEDTNKGARCSNKSFLAANBEIL 418
Db 443 SKLLLSRLDAELETRLHSQITERDETEROMKLTETLITNHHKTKTSYI----- 494
Oy 419 ESIRAKGIDIDNVKSPDETEKDNETENDSKDAEKNREEFODSLKSDSDKTPDDPEQ 478
Db 495 -----EIEQAKN-----ELLEKEVLDEDEK----- 516
Oy 479 GKSEVDGFKSEKNGELSESPAGKASGSTRITIRLRNPDLSQLKSOQVAAAAHEAN 538
Db 517 -----GDASESOSIE-----GTRKOECKKVTN-----OKSNOLT----- 547
Oy 539 KLFKEGKEVLVYNSOGELISRLSTKKEVIMKGNINNYEKLGOEGKYRVYHNOYSTNSFALN 598
Db 548 -----NGLHFKLMEGEGFNYYNOYSTNFIALN 576
Oy 599 KQHREDHDKRRHLAHKFLTPAGEFKWNGSVHGSVLTISTRLTLITOLENNIPSSFLH 658
Db 577 KQORNEERDKRRHLSHKFLTASDFKWTGTMGTDMNITTLRQTLINFEESNIAASFILN 636
Oy 659 PNMASHRAWIKAVQCSKPREPALALALIECAVKNPVYMLPIREELGTRHLRHMSIER 718
Db 637 IMVYVKKKIMNAAVNARPSSEFAVLLLPQASLKSVEFANVWHEDLGHTLORITSAER 696
Oy 719 EEKEVKKKKKKKO---EEBETMQATWVKTTPVVKHQVWKQKEEYRVYGGGSMVSKT 775
Db 697 EERKKLEKREKREDEEERNNRLAFNYIKYTLGLKHQVWKQKEEYRVHGWMLWSS 756
Oy 776 -----HYFVFKPLPGNTVYVNRKSLGTRKNMNMENMDESCKRCSRPKKIKIEPDSEK 830
Db 757 RRCGVARRAQPLTHNRVYVHYTM---GEENDVNEIIT---LVDPRTQRFMOQCES 805
Oy 831 DEYKSDAAKAGADQ--NEMDISKITEKKDQVYKELSDSDCKPEPEVDDMKTESH 888
Db 806 SNVDGQVCHYLPDQYKNVYIEDVTE-----KIKGH 836
Oy 889 VNCQESSQYDVVNVSEGFHL--RTSYKKKTKSKLDGLERIKOTPLEKORLEKIKLE 946
Db 837 -----IDVSKALNAPGRTYYSKVAARKSRDLDDLDRRLKLAIEVEQOM---ASKIP 882
Oy 947 GGIGIGIKTSTNSSKNLSES 966
Db 883 SDMKPLVSSQNNNTANSKOT 902
```

```
RESULT 9
US-60-173-464-22080
; Sequence 22080, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: C1000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22080
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
NAME/KEY: VARIANT
```


LOCATION: (1)...(786)
OTHER INFORMATION: Xaa = Any Amino Acid
US-60-173-464-22080

Query Match 10.4%; Score 1493; DB 27; Length 786;
Best Local Similarity 40.5%; Pred. No. 1.4e-78;
Matches 316; Conservative 110; Mismatches 224; Indels 130; Gaps 14;

```

QY 4 EEEEDGDAEETODSDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFSHTYSS 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 ESTHYHGSDEGDESDKSDNEDMLTPSDDESLVANESESEFVC--SFQNGV--- 131
QY 64 TPCRRPRVHRPSPILEE-KDIPPLEFPKSSDLVNPENHINVAIYEVLRNFGTVLR 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 152 ---GPRPRPSPEPVMLQEGROYALDLPDSSDDLFIANTHVALISTYEVLRFRIMVR 208
QY 123 LSPFREDFCAALYSOQCTLMAEMHVVLLKAVLREEDTSNTTFGPADLKDVSNTLYFT 182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 LSPFREDLCAALACEQSALTEVHIMLKALIRGEDAGCTHFGPLDQDYNISLYLI 268
QY 183 DGMTPEVLRYCESDKEY---HHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTTNIA 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 DSTTPEVLRYSVESDKTFRNVPFIITSHTE---YPTGIDNLEVLQFLSDQFLISNLI 325
QY 239 REELMSEGLQYDDHCRVCHKLGLDLCCTCSAVYHLECVKPLPEEVEDEMOECVVAH 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 326 RDVWLQGPFIHYDDHCRVCHRLGLDLCCTCPAVYHLECVDPMDVPTEDMOGCLCRSH 385
QY 299 KVGGVYDCAVEIQRNRYTHIEPIGYDRSRKKYWFNLRLIIEEDTPENEKKIWTYSTK 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 386 KVGGVYDCAVLPQKQVLIHSDISGVDRHGRKYWFNLRLI--EDQENFT---CWYSTT 441
QY 359 VOAEELIDCKDYWEALCKILEEMREELIHRMDITEDITLTKARGSKSFLAANEEL 418
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 SKTLKILSRDABELETIRLSQITERDELEIRQMKLETITLNEKHKTRSVI----- 493
QY 419 ESTIRAKGDIQNVKSPEETEKDKNETENDSKDAEKNREEDQSLKSDSKTPDDDEQ 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 494 -----ELIQEAKN-----ELLEKEVLIDEDKD----- 515
QY 479 GKSPVGDGFKSEKNGELSESFGAGKSGSTRILITRLRNDSKLSQKSOQVAAAEAN 538
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 516 -----GNAKSESOSTE-----GTKQOECKKATR-----OKSNOLT----- 546
QY 539 KLFKEGKEVLVNSQGEISRLSTKKEVIMKGINNYPFKLGOEGKYRVYHNOYSTNFPALN 598
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 547 -----NGTILHFKLMEGQFKYVAVQYSTNPIALN 575
QY 599 KQHREDHDKRRHLAKFCCLTPAGEEFKWSVHGSKYLITSLRLITTOLENNIPSSFLH 638
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 576 KQORNEERDKRRHLSHFSLSLTASDEFKIGITMGTTDNMTITTLRQTLINFESNIAASFNL 635
QY 659 PNMAASHANIKAVQMSKREFALALILCAVAKPVYMLPIREFLGHRLHMSIER 718
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 636 INNVVNNKINAAVMAARRSEFAVLLLEFQASLSKSYFANVWHEQGLHTTLQRTISAER 695
QY 719 EEKEKVKKKKKQ---EEETMQOATVVKYTPVKHOMVKQGEYEVYTGGSWISKT 775
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 696 EEKKLEKRRERDEENRILAFNIKTYLGLKHQWVKQGEYEVYHOMGWLMLSSS 755

```

RESULT 10
US-10-029-386-32957
Sequence 32957, Application US/10029386
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MICA-X-2
CURRENT APPLICATION NUMBER: US/10/029, 386
CURRENT FILING DATE: 2001-12-20

```

QY 67 RRRPRVHRPSPILEEKDIPPLEFPKSSDLVNPENHINVAIYEVLRNFGTVLR 126
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 RRRPRVHRPSPILEEKDIPPLEFPKSSDLVNPENHINVAIYEVLRNFGTVLR 60
QY 127 REEDFCAALYSOQCTLMAEMHVVLLKAVLREEDTSNTTFGPADLKDVSNTLYFTDGMT 186
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 REEDFCAALYSOQCTLMAEMHVVLLKAVLREEDTSNTTFGPADLKDVSNTLYFTDGMT 120
QY 187 WPEVLRYCESDKEYHHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTTNIAEELMSEG 246
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 WPEVLRYCESDKEYHHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTTNIAEELMSEG 180
QY 247 VIQYDDHCRVCHKLGLDLCCTCSAVYHLECVKPLPEEVEDEMOECVVAHKGVTDC 306
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 VIQYDDHCRVCHKLGLDLCCTCSAVYHLECVKPLPEEVEDEMOECVVAHKGVTDC 240
QY 307 VAEIQKKNRYTHIEPIGYDRSRKKYWFNLRLI 339
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 VAEIQKKNRYTHIEPIGYDRSRKKYWFNLRLI 273

```

RESULT 11
PCT-US00-05989-816
Sequence 816, Application PC/TUS0005989
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Human Pancreas and Pancreatic Cancer Associated Gene Sequences
FILE REFERENCE: PA105PCT
CURRENT APPLICATION NUMBER: PCT/US00/05989
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE

QY 2527 VAAPCPVTPVLAPAPPPSPPPPGVOHTGLSTPTLPVASOKRRREEKOSSSKSKK 2586'
Db 158 VAAPCPVTPAPAPPPSPPPPAVOHTGLSTPTLPASOKRRREEKOSSSKSKK 217
QY 2587 KKMISTTSKETKKDTKLYCICKTPTYDESK 2615
Db 218 KKMISTTSKETKKDTKLYCICKTPTYDESK 246

Search completed: November 20, 2002, 16:42:30
Job time : 337.179 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:28:06 ; Search time 19.068 Seconds
(without alignments)
4939.150 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MVSEEEEDDDAETQDSE.....KIKFKASRSHNNKLOSTAS 2781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 96903 seqs, 33865481 residues

Total number of hits satisfying chosen parameters: 96903

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents, AA, New:**
1: /cgn2_6/ptodata/2/paa/RCT_NEW_COMB.pep:**
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:**
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:**
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:**
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:**
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:**
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	432.5	3.0	2137	6	US-10-092-411A-4463 Sequence 4463, Ap
2	410	2.9	3696	6	US-10-092-411A-5080 Sequence 5080, Ap
3	401	2.8	3256	5	US-09-724-676-65945 Sequence 65945, A
4	392	2.7	2896	5	US-09-724-676-65941 Sequence 65941, A
5	390.5	2.7	4168	5	US-09-724-676-69138 Sequence 69138, A
6	390.5	2.7	4189	5	US-09-724-676-69127 Sequence 69127, A
7	388.5	2.7	3259	5	US-09-724-676-69160 Sequence 69160, A
8	385	2.7	3698	5	US-09-724-676-68931 Sequence 68931, A
9	385	2.7	3910	5	US-09-724-676-68917 Sequence 68917, A
10	380.5	2.7	3907	5	US-09-724-676-68928 Sequence 68928, A
11	380.5	2.7	3919	5	US-09-724-676-68916 Sequence 68916, A
12	379.5	2.6	3250	5	US-09-724-676-69168 Sequence 69168, A
13	379	2.6	3852	5	US-09-724-676-68911 Sequence 68911, A
14	379	2.6	3852	5	US-09-724-676-68912 Sequence 68912, A
15	379	2.6	3852	5	US-09-724-676-68913 Sequence 68913, A
16	379	2.6	3864	5	US-09-724-676-68923 Sequence 68923, A
17	379	2.6	3864	5	US-09-724-676-68924 Sequence 68924, A
18	379	2.6	3864	5	US-09-724-676-68925 Sequence 68925, A
19	379	2.6	3868	5	US-09-724-676-68914 Sequence 68914, A
20	379	2.6	3873	5	US-09-724-676-68926 Sequence 68926, A
21	377.5	2.6	4180	5	US-09-724-676-69132 Sequence 69132, A
22	374.5	2.6	3861	5	US-09-724-676-68907 Sequence 68907, A
23	374.5	2.6	3861	5	US-09-724-676-68908 Sequence 68908, A
24	374.5	2.6	3861	5	US-09-724-676-68909 Sequence 68909, A
25	374.5	2.6	3873	5	US-09-724-676-68919 Sequence 68919, A
26	374.5	2.6	3873	5	US-09-724-676-68920 Sequence 68920, A

27	374.5	2.6	3873	5	US-09-724-676-68921 Sequence 68921, A
28	374.5	2.6	3877	5	US-09-724-676-68910 Sequence 68910, A
29	374.5	2.6	3889	5	US-09-724-676-68922 Sequence 68922, A
30	372	2.6	2722	5	US-09-724-676-55868 Sequence 55868, A
31	372	2.6	2995	5	US-09-724-676-55866 Sequence 55866, A
32	372	2.6	2995	5	US-09-724-676-55867 Sequence 55867, A
33	372	2.6	3009	5	US-09-724-676-55863 Sequence 55863, A
34	372	2.6	3020	5	US-09-724-676-55864 Sequence 55864, A
35	372	2.6	3020	5	US-09-724-676-55865 Sequence 55865, A
36	372	2.6	4368	5	US-09-724-676-69121 Sequence 69121, A
37	371	2.6	4063	5	US-09-724-676-69120 Sequence 69120, A
38	369.5	2.6	2542	5	US-09-724-676-55851 Sequence 55851, A
39	369.5	2.6	2915	5	US-09-724-676-55849 Sequence 55849, A
40	369.5	2.6	2915	5	US-09-724-676-55850 Sequence 55850, A
41	369.5	2.6	2929	5	US-09-724-676-55846 Sequence 55846, A
42	369.5	2.6	2940	5	US-09-724-676-55847 Sequence 55847, A
43	369.5	2.6	2940	5	US-09-724-676-55848 Sequence 55848, A
44	369	2.6	2390	5	US-09-724-676-90223 Sequence 90223, A
45	369	2.6	2390	5	US-09-724-676-90224 Sequence 90224, A

ALIGNMENTS

```
RESULT 1
US-10-092-411A-4463
; Sequence 4463, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: 032196-101
; CURRENT FILING DATE: 2002-03-07
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-4463

Query Match      3.0%; Score 432.5; DB 6; Length 2137;
Best Local Similarity 17.2%; Pred. No. 0.00066;
Matches 395; Conservative 355; Mismatches 999; Indels 547; Gaps 77;

QY 394 ITEDLNNKARGSKMSPLAANEI-----LESTRAKKGIDNVKSPETEKD- 440
DB 149 MSSEVNTDSSSEKAGISQSSSETSNOSSELTNYASTDHESTTYNDNAQDDKSSN 208
QY 441 --KNETENDSKDAEKREFEEDSLKEDSDKTPDDPEQKSEVDFKSEKNGELSES 498
DB 209 VTKSKVQNSTSSSEKMISSNLQSIETKATDLSLATEARTSTQJNLSTSTSNSSPT 268
QY 499 PGAGKAGSAGSTRITRLRNP---DSKLSQKSOQYAAAAHEANKLF-----KEG 544
DB 269 SFANLTFESRFYVLNMAAPTSTTSTTSSLTNSVYVKNDFNEHMLNSGATYDPKKG 328
QY 545 KEVL---VNSQGEIS---RLSTKKEVINKGNIN--NYF----- 575
DB 329 IATLTPDAYSQKGAISLNLRLDSNRSEFTGKYNLNRLEGYSPDGVGTGGDIGFAFSFG 388
QY 576 ---KLQGEQ-----KYRVHNOYSTNSFALNKHQHR-----D 605
DB 389 PLGQIKKEGAANGIGLNNAFGFKLDTYHNTSTPKDAKADPRVVGGAFGAFAVSD 448
QY 606 HD---KRRHLAKFKCLTPAGE-----FKWNGSVHGSKVLITSTLRLTITOLENNIPSS 655
```

Db 449 RNMATTEASSAAKLNVOPTDNSFODVIDYNGD---TKVMVTVYAGQTFTR----- 497
 QY 656 FLHNWASHAMTKAVOMSKPREFALALILECAVKNPVMLPIMGEFLGHTRLHMTS 715
 Db 498 -----NLTWTKN---SGGTFSLMSTASTGCAKNLOVOFGTEYESAVAKVRY 545
 QY 716 IEREK-KEKYKKKEKKEEETMOATVWVYKTFPVKHQV--WKOKGEYEVY-----GY 766
 Db 546 VDANTGKDIIIPKTIAGEVATV-----IDKQNLNKNKSGVSYVSTALQMSNY 595
 QY 767 GGSMT-----SKTHVRFVFKLPGLNINWNRKSLBETKKNMDENMDSKRCRSR 818
 Db 596 SETSGPFTLKLINSQVIVIKF-----KDVQGPQISVD-----SQIREGKT 637
 QY 819 PKRKI-EPSEKD---EYKGSDAKAGDO-----NEMDITKEKDDVVELL 864
 Db 638 INPTITTTDSKDVLTFTYTGLPSCGLSPDQTTNTITGTSEVGTIVT-----VNTT 690
 QY 865 DSDSKPCKEPEMVDMDKTESHVNCQESSQV-----DVVNS 903
 Db 691 DATGNVTSKQFTTIQDTISPVVWVTPSQASEVFTPINPTITATDNGKVVTHTVGLP 750
 QY 904 EGFILRTSYKKTKSSKLDLLEKRIKQFTLEKQLEKIKLEGKIGKIGKTSNNSKNL 963
 Db 751 QGLKFDASTSIYGTPTQIGTNTITITESTDASGNKTTKINYE-----VTRNSASDS 802
 QY 964 SESPVIRKAKEGCO---SDSMROEOSPANANDQPEDLIQGSQSDSVLRMSDPGHTT 1018
 Db 803 TSTIVSVSTSIENSTSLSDSVKASOSLSTKSLSESL--SASTSNSTSIQASEASTS 860
 QY 1019 NKLYPKDRLVDVSIKSPETKCPKQNSIENDIEKVSDLASRGOEPTKSTKGNDFIDD 1078
 Db 861 KOL-----SESASTSDSASESARKSESTKSTS-----LSE 893
 QY 1079 SKLASADIGTLCKNNKPLQIESPTIVASSKALHSVPKSNDDADPLSRAMFEG 1138
 Db 894 STSTSVSDASVSTSES---ASTSTVSGSTSTSIIDSTSTSDASIKASASAS-TS 948
 QY 1139 KLGCDSNSNSTLNSSDTVSIODS-SEEDMIVONSNEISIEOFTRQDOVEVELEPKCEL 1197
 Db 949 KILSESVSTSDASASTSVSDSNASASTSLKSTSTSVSDSTSTSD----- 997
 QY 1198 VSGSTGNCEDRLPVKGTENGKKPQOKKLEKPPVKKCSDOIKLKTDTTKNNENNESE 1237
 Db 998 --SASTSTSE-----SESDSASTSLSESTSTSVSDSTS-TSTSDSASASASESE 1043
 QY 1258 KKGORTSTFOINCKDNPKIYIKGECLEKTESRVSGVNEPKVNNINKIIPENDIKSLT 1317
 Db 1044 SNKSTST-----LSF-STSTSLSGSTASST-----SDSASTS 1074
 QY 1318 VKESAIRPFINGDIVIMEDFERNKSETRKSHLSSDAEGNYRDSLETLPSTKESDSTQTT 1377
 Db 1075 TSESE-----SDSTSTSLSESTSTSLSGS-TSASTSDASST--STSESDSTSES 1120
 QY 1378 TPASACSESNVNOVEIMEFTSEVKKVTSPTTSEESN-----LSNDFIDENGL 1428
 Db 1121 T--SLSESLSTSVSDSTASSTSE---SASTSTSESESNASASTSLSGSLSTSIIDSTST 1173
 QY 1429 PINKENAVN-GESEKRVIVTEVTMTSTVATESKTVIKVEKGDQYVVSSTENCAKSTVT 1487
 Db 1174 STSDSASTSTSESDSTSTSLSESTSTSLSDSTSTSEASASTSESDSTSTSLSE 1233
 QY 1488 TTTTTVTKLSTPSTGSDVI---ISVKEQSKTVV---TTTVDLSLTPTTGGTLVMTVS 1540
 Db 1234 STSTSVSDSTSTASTSDASTSTSVSDSEASASTSISESLSTSVSDS-TSTSTSDASSTST 1292
 QY 1541 KEYSTROKVLAMKFSRKRTKRSKSTALPSYKRFVTKSKTSKSLFVLPNDLKLAKGKIRE 1600
 Db 1293 ESDSTSESTSLSESTSTSVSDSTSA--STSDASTSTSES-----ESDASTSLSG--- 1341
 QY 1601 VPYNNYNAKPLDIWPPSPRPRTGTITWRVRLQTVKSLAGVSLMLRLMLASLRMDDMAK 1660
 Db 1342 -----STSTSL-----STSTSL----- 1348

QY 1661 VPPGGGSTRPTSETTEITTTIIRKRDVGPYGRFEXCIKRIICPIGVPTPEKTEPTPOR 1720
 Db 1349 -----DSTSTSDASTSTST-----ESDSERASTSL 1375
 QY 1721 KGLRSSALRPKRPEPTKQGPV-----IETWVAEELELMEIRAFER 1764
 Db 1376 SGSTSTSDSTSTSTSDASTSTSVSDSNASTSLSGSLSTSVSDSTSTSTSDAST 1435
 QY 1765 VEKKAQAVQOAKKRLQOQKPTVIATSTSPISSTSTISPAQKVMVAPISGSVTTGTR 1824
 Db 1436 SEDD-----SERASTSLSGSTSTSIIDSTSTSDASSTVSE-----SNSTSTSL 1484
 QY 1825 MVLTTKVSPPATVTPQONKNEHOTFATVWKGOSNSGVVOQKVLGIIPSTSTGTSQOFT 1884
 Db 1485 ESLSSTSV-SDSTST-----STSDASTSTSVSDSDASTSSSSSV-----STSDSEST 1532
 QY 1885 TSFOPTAIVT-----RPTSGSGGTTSSQVITGQPIRPGMTVIRTPLOQSTLGKAIT 1939
 Db 1533 TSTSDASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSDASTS-----ASTSESDAST 1589
 QY 1940 RFPVMVQPGAPQOVMTQILINGQPVSTAVSAPNTVSSNPGOK---SLTSASTSTSIQSSAS 1996
 Db 1650 STSESDASTSLSDSTSTSVSESTSTSTSVSASNSTSTSLSD--DSRSTSLSDSTST 1707
 QY 2057 LMOQAMPNGIVQRLPPLATATATSTTTTYYSTTRAGCEGROKSLSTQOMVHODKTL 2116
 Db 1708 STSESGSTSTSES--DSDASTSLSESTSTSIIDSTSTSDASTSTSVSDSNASTSL 1765
 QY 2117 PPAOSSSV--GPAKAPQOTAPAPAPQOPQOPQAPQOPQOTQOTVTVSSHVPEA 2174
 Db 1766 SDSTSTSVSDSTSTASTSEASTSTRESESTASTSLSESTSTSVSDSTSTSD---SAS 1822
 QY 2175 QPTHAOSKRPQVAAQOSQVNOGQSPVNOGQSPQTRIRPSTPQSLSPGOQSOVQTTTQ 2234
 Db 1823 TSTSESDSNESSTSLSESTSTSTSDSTSTASTSTSVSDSNASTSLSGST 1877
 QY 2235 PIPQIPTSLQIPQOQOPQOPQOVOSTOTLSSGQTLNOVSVSPSRPQLOIQOPQOVT 2294
 Db 1878 STSVSDSTSTSTASASTSTSESDSDASTSLSGSTSTSIIDSTSTSDASTSTSE-- 1935
 QY 2295 AVPOLQOQOVVLQIOQSOVVAQIOAQOSGVPOQIKLOLPTIOIOSSAVQTHQIONVTVQ 2354
 Db 1936 -----SASTSTSVSESDSTSESTSVSESS-----TSVSDSTSTSTSEASTSTSE 1979
 QY 2355 AASYOEOLOQVQOLRDOQOKKKQOQOIEIKREHPLQASNOSEITIQOVYMKHNAVIEHLK 2414
 Db 1980 SESTSES-----TSVSESSSTSIIDSSSTSTSMSTSEFTTQSPINSSEOFIDGSLSE 2032
 QY 2415 KKSSTPAREBNORMIYONVMYILDKIKDEKQAKKKRRESVQOKSKONATKLSA 2474
 Db 2033 DTIVT---QSKNTNML--NKTGK---DYDLOQKGTDEOHNEHQSNQADNHNNMLD 2083
 QY 2475 LFFKKEQLRAELIKK 2490
 Db 2084 L--HQRNRLODKVVKQ 2096

RESULT 2
 US-10-092-411A-5080
 ; Sequence 5080, Application US/10092411A
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/10/092,411A
 ; CURRENT FILING DATE: 2002-03-07
 ; PRIOR APPLICATION NUMBER: US 09/134,001

```

; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-5080

Query Match      2.9%; Score 410; DB 6; Length 3696;
Best Local Similarity 17.7%; Pred. No. 0.0034;
Matches 464; Conservative 427; Mismatches 1027; Indels 704; Gaps 108;

QY 311 OKKKPIRHEPIGYD-RSRKKWFLNRR--LIEEDTENENKKIYVSTKVLAEIDLC 367
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 569 QYKKDI---PSNYTLASTNKYIKLKERAQTVLDEETNTP----- 606

QY 368 LDKYWEAECLILEMREIEHRMDITEDLTNKKARG-----SNKSFLAANEELLES1 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 607 FNGYSOTQIDDLHELQTLINRVSASREINDKQEMTDVAYDSDELTEEKDTLVQI 666

QY 422 RAKKGIDVAKSPE-----ETEKDNLETENDSKDAEKNREPEDOGLE 464
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 667 ENKNKISNNIDDELTDGVERVKEAGLHLESDTPHPVTKPMAROVVNNRRAD-QCKTLI 725

QY 465 KOSDDKTPDDPE-----QKSEVGDPEKSEKSNGLSPGACGASGSTRITTR 514
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 726 RNNHEATTEONAIROVEAHSSDAIAKIGEAETDTTVNARD-----NCKTLATD 777

QY 515 LRNPDSKLSQSOQVAAAHAEANKLFKEGKEVLVNSOGESIRLS-----TKKEVI 566
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 778 VNPPTK-----KAEARAATVNSANSKIKD---INNNTQATLDERDALALVNRKDEAI 828

QY 567 MKGNINNYRKLGOEGKYRYVHNOYSTNSF-----ALNKHOREDHDKRRHLAH 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 829 Q--NINT--AOGNDVTEAQN--GTNTIQOVLPTPEVKRONALATINAKADEQKRLIQAN 882

QY 615 KCLTPA---GEFKMGV--HGSKVLTISLRLITOLENNIPSSPLHPMAHRAHWIK 670
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 883 NNATTEKDAEKKVNAVITANONITNATNDVDAQTT-----GSGTIS 929

QY 671 AVOMCSREFPALALILECAVAVVLMPIWREFLGHTLRMTSIEREKEKYKKKKKK 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 930 AISPATRIKEDARA-AVEKKAIAQNOOI-----NSNNATTEBEKDALNOVEAKH 978

QY 731 QEEBELMQOATVWKYTFPVKHOVMKOKGEBYRVTVGYSWISKTHVYRFPKLPQNTNY 790
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 979 QAAIATINOA-----OSTQOVS-----AKNNGINTI 1005

QY 791 NVRKSLGTRKNMNDENMDESDKRSPKIKIIEPDSEKDEYKSGDAKGAQONEMDIS 850
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1006 NODQPAVAKNNKTKTLEQNGNEKS---AIAQTPDATEE-----KOEAVS 1049

QY 851 KITEKKDQVKELLDSQPCKEEPEVDDDKTESHVQCOSSQSDVYVSEGHART 910
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1050 AVSQATNIGITHNOANSNDVDQE--LSNAEQITHTNVAVQKKPQ-----AQALIAKT 1103

QY 911 SYKKTKSSKLDGLERIKQF-TLEEKORL-EKI-----KLEGIGIGKTS 956
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1104 NEKOSAINDNDEGTIEEKOAIOSLNDAKNLADEOITQAASNONVNALNIGISNISKI 1163

QY 957 TNSSKLSSPVITAKKECCQSDSMQEDSPANNDQPPDLLOGCSOSSSVLRMSDPIS 1016
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1164 TNEFK---KQARDOVNOQFQEKAEALNSTPHATODEKODALTRLQOAKETALINDINOQ 1220

QY 1017 TTNKLYPKORVL-----DDVSIKSEPTKCPKONSIEINDIEEKVSLASGQEPKTS 1067
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1221 TNOVN---DTALTSQIONTONOVNVR---KQEAKTITINDIQOIKKOSTIQNNDATTEE 1274

QY 1068 KTKGNDFFIDSKLASADIDIGLICKNKPLIOESDITVSSSKSALHSSVPKSTNDRDA 1127
```

```

DB 1275 KEVANN-----LVNASQ-QANVISKIDNATNNQIDGIYSDRGOSINATPTPTSIKRNA 1326
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1128 TPUSRAMDEFGKLCGSESENSTLENSSDVTIO---DSSEDMATVONSNSISEQRTRE 1384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1327 -----KNDIDIKADKKIKIORINDADEE--IOPANKTIE----- 1361

QY 1185 QDVEVLEPLKCELVSGESTGNCEDRLPVKGTENGK---KPSQOKKLEERPV--NKCS 1238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1362 -----AKIEAKDNIIQRNSTRDQVNEAKTNGINKIENTIPATVTKSEARQAVQKKANE 1413

QY 1239 QIK-LKNTTDKKNNENRESKKQGTST-----FOINGKDKPKI-YLKGELKEIS-- 1288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1414 QINHONTDPATNEEKEOEAL--NRVSAELARVOAQINMEHTQGVKTKIDDAITSLSR 1470

QY 1289 -----ESRVSGNVEPKVNNINKIIPENDIKSLTVESAIRPFINQVIMEDENRNSSE 1343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1471 NAQVEKESARNAIEOKATQOQTFINNNDNATDEKEVA-----NNLVIA--TKOKSD 1522

QY 1344 TKSHLSSSDAE---GNVRDSLETLPSTKESDSTQTTTPSASCPSNSVNOVEDMEIET 1399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1523 NINSLSSNDNENAKVAGIEIANVLPATPAVSKAKKIDQKL--AQQINQIQHTQAT 1579

QY 1400 SEYKATVSSPTTEESNSLNDPIDENGLEPKNNENVNGESKAKKYITTEVTTFTSVATE 1459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1580 TEEK-----EAAIOLANOKSNEARTAIQNEHNSNGVAQAK----- 1614

QY 1460 SKTVIVE-----KGOKQTVSSNENCAKSTVTT-----TTVTKLSTPSTGS 1503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1615 SNGIHETELVMPDAHKSDAKOSIDMKYNGQSNITNTPATDEEOKALDKLIKADNG 1674

QY 1504 --SVDIISVKEQ--SKTVTTVTVDLSLTGGLTVSMTSVSKYSTRPVKL--MKFSRP 1557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1675 YKNVDQAQTMQOVSADKATEADITITN-----IQANVAKKPAR--VELDSKFEED 1722

QY 1558 KTRSGTALPSTYKKEVYKSTKSI FVLPN--DCLKKLARKG-----GIREVP 1602
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1723 K--ROIATAPN--ATEEKKODAIORLNGKRDENVNLINODRDNDEVEOKNIGIQELE 1776

QY 1603 YFNYNKAPALDIPRPSRPTFGITWRYRLQTKSLAGVSLMLRLMASLRMDMAKVP 1662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1777 TIHAN-----PTKSDALQELQF-KFISQELI----- 1803

QY 1663 PGCGSTRTESETEITTTTELIKRRDVGPGYIREYCIRKIIICPIGVEPPEKPTPORG 1722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1804 ---NNKKAIDNEKDEAKRL-----EISKNTITINQAQTMQOVDNKKDNG 1848

QY 1723 LRSSALRPKRPEPTKOTGPVYIETWVAEELELMETRAFAERYEKEKAQAVEQAKKRL 1782
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1849 MNEIA--TIIPATYITDAKTAIDKRAEQOVTI--INGNNDATDEKAEARKLVERAKIE 1904

QY 1783 Q-----OKPTVIATSTSPSTSTSTISPAQKVNV-----APISG 1817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1905 AKSNTINSDTEREVNGAKTNGLEKINNIOPTQTKTNAQOETINDKAOEOLIQINNP-- 1961

QY 1818 SVTGTAKMVLTVKVGSPATVTFQONKNFHQTEATWVKQGSNSGVVOVOQKVLGIIPST 1877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1962 DATEEKEQKATNNVNMGLAQAIQINNAHST--QEVNESKTSN--IATKISVGNVIAKKPT 2018

QY 1878 GTSQOFTFSFOPT-----AT-----VTF-----IIPNTSGSGGTSNSQ 1911
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2019 AINSLTQOANNOKTTLIGDNGNATDEKEAKAKOLVYOTKLNEOLOKIHESSTODNOVDNWKQ 2078

QY 1912 VITGPQI-----RPGMVIPTPLDOOSTLGKALITTPVWVGCAQOQVWTOIIRQVP- 1963
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2079 AITAIKLINNAHAKKRODAINILITNLAESK--KSDIHAN--ODATTEENTAI--QSD 2130

QY 1964 STAVSAPNTVSSPTGOKSLTSATSTNIOSSAQPPR-----POGQOVLTMAQLTOLQ 2018
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2131 DTLAQARRNING-----ANTMNLVYDENLEDGKOKLORIVLSTQTKQAKADIAQAL---- 2181

QY 2019 GHGNGGLTVVIOGQGTGQLOLIPQGVTVLPGEQQLMQAAMPNGTVQRELFPLPAT 2078
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


Dd	2182	---	GOORSTIIDONONATTEKEKOEALERLNOETNGVNDRIQALANONVTPDEKNNILETI	2257
Qy	2079	ATT-----	ASTTTTTTSTAAGTGEOROSKLSFOMOVHODKTLPPAOS	2121
Dd	2238	RNVPIVIVPKRAEIRIRKKAABOTLLINONODATLEEKQIOMLGKLEFVKKEALNVSQA		2297
Qy	2122	SSVGPAAQAPOTAPARPOPTOQOSAPQEVQ--TQPEVQT-----		2162
Dd	2298	HSNNDVIAENNGIATKISEVHPETIIKRNKQOEIDQASQJDTINANKNSTNEKSAI		2357
Qy	2163	-----	QTTSSHVSEAPPHAAOS-----	2182
Dd	2358	DRVNAKIDAINNTTNAATTQVLVNDAKNSGNTSISQILPSTAVKTNALALASEAKKNNA		2417
Qy	2183	---KPOVAASQPOSN-----	---YOGSPYRVOSPS	2208
Dd	2418	IIDQTPNATLEEKKEANKNVDRLQEEADANILKHAFTTDEVNNIKMQAVININAVQEVIK		2477
Qy	2209	QTRIRPSTPSQLSGQOSQVQTTSSQPIRPHPTISQITPSQSQPOPSQOVQSSVOTLSSG		2268
Dd	2478	KONK-NOLNOFIDNOKRIENT-----	PDATLEEKAEANRLLQNVLTSTDEIANTV	2528
Qy	2269	QTLNQSVS--SPSRPQIQIQPOPOVIAYVOLQOQVQVLSQISQOVVAQ--IQAOQSGV		2324
Dd	2529	DHNNEVDQALDKAPKPIE-----	EIVQVSKKQDVNALIQEAFNSQIQEIQONQAT	2580
Qy	2325	POQIKLQLPQIQOQ--SSAVQHQIQONVVYVQAASVQEQQLRVQQLRQOQKKQOQIE		2381
Dd	2581	NEE-KTEALINKINOLLQNAQVYNIDQASNKEDVDSAKTR-SIDIEIQPHQPTK-----A		2633
Qy	2382	IKREHTLOASQMSLIIQOVYMMH-NAVIEHLKQKSMTPARREENGRIYONOVMYIL		2440
Dd	2634	TGRHRLNKKAQO---QOSTTATHPNSTIE-----	EKEQSAKL-QEVLKRAI	2676
Qy	2441	DKIDEEKOAKKRRKRESEYEOKRSKONATKLSALLFKHEQLRAEILKRRALLDKDQOI		2500
Dd	2677	AKIDKQGVNDVQKTYVNGIAIEINILPATTV-----	KDKAKADVNAKE-QKNQOI	2727
Qy	2501	EYQEEKFKDKLIKKEKDIMOLAQATAVADPCPVTVPLDAPPAPEPPPGVOHTGLL		2560
Dd	2728	NSNDEATTEEKLTVASMDINLHVETTNQAIIDADDTQOVNVE-----	KNKGIG	2774
Qy	2561	SHPTLPVASQRRKREKEDSSKSKKMMJSTSKETKDKTK		2602
Dd	2775	TIRDIQPLVAKRPYAKSKIESAVEKKEKTEITQONATHDEVR		2816
RESULT 3				
US-09-724-676-65945				
Sequence 65945, Application US/09724676				
GENERAL INFORMATION:				
APPLICANT: Comugen LTD				
TITLE OF INVENTION: Variants of alternative splicing				
FILE REFERENCE: 129181.4 Comugen				
CURRENT APPLICATION NUMBER: US/09-724, 676				
CURRENT FILING DATE: 2000-11-28				
NUMBER OF SEQ ID NOS: 97222				
SOFTWARE: PatentIn version 3.2				
SEQ ID NO 65945				
LENGTH: 3256				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-09-724-676-65945				

```
Query Match      2.8%   Score 401; DB: 5; Length 3256;  
Best Local Similarity 18.7%; Pred No. 0.005;  
Matches 557; Conservative 364; Mismatches 1016; Indels 912; Gaps 131.  
  
QY    339 IIBEDTENENEKKIWMYSKRVOLAEILIDCLD---DYWEALCKLIEEMREIRHRMDI 394  
       ::::|||||::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     432 VLEETITENNEPPLTLMILQVERKIORDSLSPKEIKGTAGGCSGLPELS-----VDI 486  
       ::::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
QY    395 TE-DLTNRARG-----SNKSFLAANEELTE-----SIRAKGDIDN-----VKSPEED 436
```

Db	487	NNFDGSDINESGJIFLKRKRRVSPFGHILRPBEPBENLPBNTPLKRGAPTRKRSKLYMHHPV	546
Qy	437	TEKOKNETENDSKDAENR-----EFEDQSL-----EKSDQKTP--DQDEQOKSEVGEFKS	488
Db	547	LKTIKEPOPOSGOGSGSGSIHVEVAQSLVISPAPSRKTPPVAASDQRRSCCTAPASS	606
Qy	489	EKSNGELSESPGACKGASGSETRITFLRMBDS-----KLSQJGSOQVAANAHAHLFK	542
Db	607	SKSQTEVPK-----RGERVATCLQKRVSISRGQHDILQJCSKRSGASBAN-----	654
Qy	543	EGKEVLVYNSGGEISRLSTK--EVIIMG--NINNFELQOEQKYV--YHQYSTN-593	
Db	655	-----LIYAKSWADVYKLGAKQOTQYTKLHQBOPKSMKROBPATPKKVEGVHISOFGSTGH	710
Qy	594	-----SPALNKHQHRREDHDKR--HLAHKFCULPAGEEK-----WNSVYGSVLYT	637
Db	711	ANSPTIIGIAHTEKHAVPAPRYRVLNNTISNQKMDFFEDISGIAEMFKTPVQEOPOLT	770
Qy	638	ISTLELTITQOENNIPSSF-----LHPNASHRAMIKAVOMCSK--PREFALALAI	687
Db	771	-STCHIAISSENILGKQFOGDTSGEELPTPSESQGVVFPASANAQKQPSD-----	822
Qy	688	LECAKPVYMLPIRRELGH-----RLHMTSIEEE-----	719
Db	823	-KCSASP-----PLRQOQIRENGVAVAKTPNTYIAIMSLETKTSDTEPESKYSTVYNSGR	877
Qy	720	--EKEKVKK--KEKQOEEMT-----QOATVVKYTFPVKHQVQKQGEERYVGY	766
Db	878	STEFENIQLKLVESKSEBTMTVEICILKRGKATLQO-----RREGEMKEI---	925
Qy	767	QGMWISIKTHYRVPKLPQNTVNVYKSLGCKNNMDENMEDSK-----RKCS--RSP	819
Db	926	-----BRPEET--KXENIELKEN--DEKMAKMRSTQWCKAPMSDL	964
Qy	820	KKIKIEPDE--KDEVKG-----SDAKGADQENEMDIKTEKKDQDYKELDSDSDKP	871
Db	965	TDKLSLPTELKMTQTAGQNLQOTQDHAKRKSEKQITIKM-----P	1006
Qy	872	CKE-EPMEVDMDKTESHVNCQSSQVDVYVNSE-----GFHLRIS-----	911
Db	1007	COSLOP-----SPIMPHYTKQQLKASIGKGVGVEELLAVGKFTPTSGTTHHREPADG	1062
Qy	912	-----YKKTKT-----SKLDGLLERTIKOFTLEEKORLEKIKLEGJIGIKTSTNSK	961
Db	1063	KIRIFKFSKPOILDPARAVTGM--KWPPTPBEHQSLD---AGRFELFQTPGPEE	1117
Qy	962	NJSESPVITTKAREGQSDSMRQESPMANNDQPE-DLIGQCSQSDSVLRMSDPSHHTNK	1020
Db	1118	SWTDEKT--TKIACKSPPEPVDTPTSTQWPKRSIRKADVDEEFILAKRLTLPBAGKAM	1174
Qy	1021	LYPKRVLVDVSI-----RSEPKCKPQKNSIEN-----DIE	1051
Db	1175	LTPKAGGDEQDIKAFMGTPOVKLIDLAGTLPGSKROLQTPKEAQALDELJAFKELFOTP	1234
Qy	1052	EKVSPLASRGQ-----BPTKSKTKGNDFFIDSKLASADDIGTLI--CKNK	1095
Db	1235	GHTHEELVAAGKTKRIPCQSDPQSDVPDPTSTQKQ-----PKRSIRKADVEGELLACRNL	1288
Qy	1096	KPLIOESDITVSSSKALMSHYKSTINDD-----ATPLSRAMPFEKGLGCDSESNTL	1150
Db	1289	MP-----SAGKAMHPKPSVGEEDIIIFVGTPOVKL-----DLT	1323
Qy	1151	ENSSDPTVSIQSSSEEDMIVQNSNESISEORPTREODVEYLEPLKCLVIGESTG--NCEDR	1209
Db	1324	ENULTGSKRRPQTPKPEAOALEDLTGFELEFQTOGHHEA-----VAAGKTKMKPCSS	1376
Qy	1210	LPVQKTEANGKRRPSQOKLLEERPYNKCSQDIKLKNTDKKNNEBSESEKKGORTFOIN	1269
Db	1377	PESADDTPTSTRROQKTPLEKRDVOQKELSLK-----KLTOGTSGETTHTTDXP	1424
Qy	1270	GQDNKPKYILGECLEKTESBRSRVYSGNV--EPKYANNINKIIPENDIKSLTVKRSALRPIN	1328

Db 1425 GGEDKSINAFRETAOKIDLPASAVTGSKRHPKTK--EKAQPLEDLAGW--KELFOTP--- 1477
Qy 1329 GDVIMEDENERNSSSETKSHLLSSSDAEGNYRDSLETLPSTKESDSYOTTPPASCPESNS 1388
Db 1478 ---VCTDKPTTTEKTTKICRCSQPD-----PVDTPITSSKQSKR 1513
Qy 1389 VNQVDEMELETSEVKKVSS-----PITSEESNLSNDFIDENGPIPK--NEWV 1436
Db 1514 SLRKVDVEEFALKRRPSPACKAMHTPKPAVSGEKNIYA--FM--GTPVOKLDTJENL 1568
Qy 1437 NGESEKRYVITEVTMTSTVA-----TESKVIKVEKGDKQVNVSTECARS 1484
Db 1569 TESKRRLOTPKKAQALIEDLAFKELFOTRGHTESMT--NDKTAKVACKSSQPDLDKN 1625
Qy 1485 TTTTITTTTUTKLTSTPSTGSDVDISVKEOSKTVTTTVDSTLTGGTLVTSMTVSKES 1544
Db 1626 PASSRRLKLTSG--KVQKEELAVGKLTQTSGETHTHTTEPTGCGKMAKPMESPKOI 1683
Qy 1545 TRDKVYLMKFSPPKTRSGTA-----LPSTYRK--FYTKSTKKSIFVLPLND 1588
Db 1684 LDSAASLTGSKQRLTPKPKSEVPEDLAGFIELFOTPSHTKESMTNEKTTKVSYRASQPD 1743
Qy 1589 L-----KTLARKGIREVPEYFNNAKPAIDIMYPSRPRTFGITWYKRLQTVAS 1637
Db 1744 LVDTPTSKPKPQPKRSLRKADTEE-----EFLAFKQTPSAG---KAMHTPKP 1787
Qy 1638 LAGVSLMLR-LIMASLRMDMAKAVPBGSGSTSTSETEITTELIRKRDVGPYGR-- 1694
Db 1788 AVGEKRDINTFLTGYOKLDOPGNLP--GSNRLOTKRKAQALELIT-----GFRLL 1838
Qy 1695 -----FEYCIRKTIIC--PIGVP--ETPKETPTPQKGLRSS-----ALRKP 1733
Db 1839 FOTPCDTNPTADEKTTKTLCKSPQSDPADTPTNTKQRPKRSILKADVEEFLAFKRLP 1898
Qy 1734 -----ETPKQNGPIVIEIWAKEEELMEIRAF--AEENE-----KE 1768
Db 1899 SAKGMAMHTPKAA-----VEEEK--DINTFVGTPEVKLIDLGNLPGSKRRPOTPRE 1946
Qy 1769 KAAAVEOQAKKRLEEOQKPTVIATSTTSPSTSTISPAQKTMVAPISGSVTTGTMVLT 1828
Db 1947 KKALEDLAGFELFOTRGHTESMTDOKITEVSCSKSPQPDVKNP-----TSSQRLK 2000
Qy 1829 TKVGSBATVTFQOQKNFHOTFATWVKOGOSNGVVOQOKVL--GIIPSTGTSOOTFTS 1886
Db 2001 ISLGR-----GVKKEEVLVPVKLQTSKTIQT--- 2028
Qy 1887 PQPRATVYIIRNTSGSGT-----TNSQVYI-----TGPOIRGMVYIRP-----L 1929
Db 2029 -----HRETAGDGCSIAFKESAKQMLDPANYGTMGERMP-----RTPKEAOSL 2073
Qy 1930 OOSTLGKAIIRTPVWVQPAQVWTOILRGQVSTAVSAPMTVSTPGOKSLTSATSTS 1989
Db 2074 EDLAGFELFOTRGHTESMTDOKIT-----ACKSPPEEMDPTST- 2117
Qy 1990 NIOSASOPRPOGOVOKLT--MAQLTDLTOGHGNOGLTVVIOGOGOTTGQLDLP--- 2044
Db 2118 -----RRRPKPTLCKRDIIVEELSAKOLTO-----THTDKVPGBE 2153
Qy 2045 -OGVTVLGRPGQOLM-----QAAMPNGTVQ-----RFLF--TPLATATTTAS 2083
Db 2154 DGIWVIFRETAQOKIDLPASAVTGSKRQRPTRPKGAQPLEDLAGLKELFOTPRCKPTTH 2213
Qy 2084 TTTTIVSTTA-----AGTGE--QROSKLS--POMQVHODKTLTPAOSSSVGPAAKQPTA- 2134
Db 2214 EKTTHIACRSQPDVGVGPTIFPKPQSKRSILKADVEEELARKRTPTSGKAMDTPKPAAG 2273
Qy 2135 -----OPSARPOPTOPQSPAPQEVQTOPEVQTOQTTVSS-----HVP 2171
Db 2274 GDEKDMAKFMGTPTYOKLPLPGMLPGSKRMPQPKR-----KAAQLEDLAGFELFCRP 2326
Qy 2172 SEAOPTHNASSKPOVAAASOPNSNVQOSPVAVVQSPQSTRIRPS----- 2215
Db 2327 GTDKFT--TDEKTTKTACKS-PO-----PDPVDTPASTKQRPKRNLKRAKADVEEFLALR 2377

Qy 2216 --TPS-----QLSPGQOSOVOTTTTSOPIPIOPHSL-OIP-SQGQOSQP----- 2256
Db 2378 KRTPSAGKAMDTPKPAVSPDEKNINFEVTPVOKLIDLGLMLPGSKRQRPQPKKAALEDL 2437
Qy 2257 -----QVQSTQTLLSSQTLNQVSVSSPRQIQOPQOVLAVP-----QLQOQVY 2305
Db 2438 VGFKELFOTRGHTESMTDOKITEVSCSKSPOPEFSFTSSSKQRLKPLVKVDMKEEPLA 2497
Qy 2306 LSOI--QSQVVAOIQAOQSGVPOQIKL--QLPIQIOSSAVOTHOIQNVV--VQAAVQE 2360
Db 2498 VSKLRTTSETTQTPTEPIGDSKSIKAFKESPKQIIDLPAASVYGSKRQRLTRKRAALE 2557
Qy 2361 QLORVQOL-----RDQOQK-----KKOOOIEIKR--- 2384
Db 2558 DLVDFKELLSAGHTEESMTIDKNTKIPCKSPPELTDTATSTKRCPKTRPKRYVEELS 2617
Qy 2385 --EHTLOASNOSEIIOQVVMHNAVIEHLQO--KKSMPAREENQNMIVGNQVKYTL 2440
Db 2618 AVERLTQTSQSTHTHKEPASGDEG--IKVLQORAKKRPMPVEEERSR----- 2664
Qy 2441 DKIDKEKQAAKKRKEEVEQKRSQNTKLSALLFKHEQLRA-----EI 2487
Db 2665 -----PRAPKKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPELV 2713
Qy 2488 LKKRALDLQDLEVOEELKRLKIKKEKDLQLOAQ----- 2523
Db 2714 VDTTASTKRHLTRQK-----VQVKEEPSAVKFTQTSGETTDADKAPAGEDKIGIKALKE 2768
Qy 2524 -ATAVAAPCPVTPVLAPAPPPSP-----PPRGVQHTGLSTPTPLVASOKRR 2574
Db 2769 SAKQTPAPASVYTGSRRRAPRASAQALIEDLAGFKDPAAGHTEESMTDOKTTKIPCKSS 2828
Qy 2575 EEKQSSSKSKK-----KMISTSETKKDKTLVYICTPTDESK 2615
Db 2829 PELEDATISSKRRPRTRAQKVEKKEELAVGKLTQTSGETTHFD-----KEPVGEK 2880

RESULT 4
US-09-724-676-65941
; Sequence 65941, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 65941
; LENGTH: 2896
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-65941

Query Match 2.7%; Score 392; DB 5; Length 2896;
Best Local Similarity 18.5%; Pred. NO. 0.0073;
Matches 533; Conservative 370; Mismatches 1025; Indels 948; Gaps 133;

Qy 291 QCEVCVAHNVPGVTDCAVEIQKNKPYIRH-----EPITG-----YDR 326
Db 34 ECDIRIQLPVVSQKQKLEIHQDEALHNFSSNTPTQVNGSVIDEPVRLKHGDVITIIDR 93
Qy 327 SRKRWFLNRLRIEDPTENENEKKIWTSTVQVQALDIDCDKQVWEALCKILEENME 386
Db 94 SFR-----YENESLQNGRKSSTEPKRIK-----DQEPARRVS--RS 127
Qy 387 EIHHRMDITEDLTNRKAGSNKSFLLAANEILE-----SIRAKGDIDN-----VKSPE 435
Db 128 SFSQDPDESEGIPLKRR--RVSFGGHLRLELDENLPRPTPLAKREAPTKRKSLYMHNP 185
Qy 436 ETEKDKNETENDSKAEKNRE--EFEDQSL-----EKSDQKTP--DDPEQKSEVGFK 487

Db 186 VLKRIIEQPQPSGKQSSGEIHEVEKAQSLVISPAPSPRKTTPVASDQRRRSCTAPAS 245
QY 488 SESNGELSSPGAGCAGSSTRITRLRMPDS-----KLSQKSOQVAAAHENAKLE 541
Db 246 SSKSQTEVPR-----RGERVAICLOKRVSISRSQHDILQMCKSRSSGASEAN-- 294
QY 542 KEQKEVIVVNSOGESISLSTRK---EYIMKG---NINNYFKLGOEGKYRV--YHNOYSTN 593
Db 295 -----LIVAKSMADVYKLGAKQOTKTKHGPORSAMKRRRATPRKPKVGEVHSQFSTG 349
QY 594 -----SFALNKHQHRDHDKRR--HLAHRFCUPLPAGEFK-----WNSVHSGSKVL 636
Db 350 HANSPTIIGIKATEKTEVHPARPYRLNFINISQKMFEDLSGLAEMKTEKTPVKEQPOL 409
QY 637 TISLRLITIOLENNIPSSP-----LHPMASHARAWIKAYOMCSK--PREFALALA 686
Db 410 T-STCHTAISSENLLKQFOGTDGSEBPLPTSESFGVNFSAQNAKQPSD----- 462
QY 687 ILECAVPVMLPIWREFLHT-----RLHMTSIERE----- 719
Db 463 --KCSASP---PLRRCIENGNVAKTPRNNTYKMTSLETKTSDTEPEPKTVSTVNRSG 516
QY 720 ---EKEVYKR--KEKQOEBEETM-----QOATWVKTFPPYKHQVWKQKEEYRVTG 765
Db 517 RSTEFRIIOKLPIVBSKSEETWTELVCEILKRGOKATLLOQ-----RREGEMKEI-- 565
QY 766 YGSMWSKTHVYRFVFKLQNTVNVYRKSLSEGTKNMNMDEMDESDK-----RKCS--RS 818
Db 566 -----ERPEET--YKENIELKEN--DEKMKAMKSRKWGCKCAPMSD 603
QY 819 PKKIKIEPDS--KDEYK-----SDAANGADONEMDISKITEKKDOVYKELIDSDSK 870
Db 604 LTDLSLPDTELMKOTARGQNLQTOHAKAPKSEKGIITKM----- 645
QY 871 PCKE--EPENVDMDKTSHVNCQSSQVDVYNVSE-----GFHLRIS----- 911
Db 646 PCOSILOP---EPINTPLTHKQOLKASLQKGVKVEELLAVGKTRTGETTHHREBPAGD 701
QY 912 -----YKKTTR-----SKLIDGLERRIKOFTLEEKORLEKIKLESGIKGTSTNSS 960
Db 702 GKSTRTFESPKOILLDPAARNTGM--KKWPTPKBEAQSLLEDL--AGFKFLPOTPEPSE 756
QY 961 KNLSESVYTKAKBGCOSDKROESPANNDQPE--DLIQCSOSDSSVLRMSDPSTHTN 1019
Db 757 ESMNDEKT---TKIACKSPPEVSVDTPSTKQMPKRSRLKRDVDEEFLAKRLKPLPSAKA 813
QY 1020 KLYPKDRVLDOVSI-----RSPETKCPKQNSIEN-----DI 1050
Db 814 MLTPRPAGGDEKDIKAENKTPVQKLDLAGLPGSKROLQTPKEKAQALEDLAGEKLEFOT 873
QY 1051 EEKVSDLASRQO-----EPTKSKTKGNDFFIDSKLASADDIGTLI--CKN 1094
Db 874 PGHHEELVAAKTKTKIPDSDPOSDPVDPTSTKQK-----PKRSIRKADVEGELLACRN 927
QY 1095 KKLPIQESDITVSSSKSAHLSVPKSTNDPD-----ATPLSRAMDEFGKLGCSSESNST 1149
Db 928 LMP-----SAGKAMHTPKPVSVEEKDIIIFVGTPOKL-----DL 962
QY 1150 LENSDDTVSIDDSEEDMIIYONSNEJISQOPRTREODVEVLEPLKCELVSGESTG--NCD 1208
Db 963 TENLTGSKRRQOTPEKEQALEDLTGKELFOTPGHTEEA-----VAAKTKMKCES 1015
QY 1209 RLVPYKGTBANGKPSQOKLEERPVYKNCSDQ IKLNTTDDKNNENRESEKKGQRTSTFOI 1268
Db 1016 SPPEBADOPTSTRRQPKPLERKRDYQKELSAK-----KILQTSGETTHHDVKY 1063
QY 1269 NGKDNKPKIYIAKGBCLAKTSSRYVSGNV--EPKYANNIKIILPENDIKSLTYKESAIRFI 1327
Db 1064 PGGEKKSINARETAKOKLDPAASVYTGSKRHPKTK--EKAOPLEDLAGW--KELFQRP-- 1117
QY 1328 NGDVIMEDEFERNSESTSHLSSSDAGNRYRDSLETLPSTKESDSQOTTPPASCPESN 1387
Db 1118 -----VCTDKPPTHHEKTTKIACRSQPD-----PVDIPTSSKPSQK 1152

QY 1388 SYNQVDEMEIETSEVKKYTS-----PTTSEESNLSMDFDENGPIPNK---NEN 1435
Db 1153 RSLRKADVEEFPFALLKRRPPSAGKAMHTPKPAVSGKNIIYA--FM---GTPOKLDLTEN 1207
QY 1436 VNGESKRKTVTEVTMTSTVA-----TESKTVIYVEKQDKOTVYSTENCAK 1483
Db 1208 LFGSKRRLQTPKEKAQALEDLAGEFKELFOTRGHTEESMT---NDKTKAVACKSOPDLDK 1264
QY 1484 STYTITTTTVTKISTPSTGSDVILISVKEOSKTVYTTVTDLSITTTGGTLVSMYKSEY 1543
Db 1265 NPASSKRRLKTSLG--KVQKEELLAVGKLITQSGTTHHEPTIDGSGMAKAPMESPOQ 1322
QY 1544 STRDYKVLKMFSPKRTKRSQTA-----LPSYRK--EYTKSTKKSIFYLPND 1587
Db 1323 ILDSASLGSKRQOLRTPKGSSEVPEDLAGFLEFOTPSHTKESMTNEKTKTVSYASAP 1382
QY 1588 DL-----KTLARKGITEVYFYNNYNAKRALDIMPSPRPRTGILTRYRLQTVK 1636
Db 1383 DLVDTPTSSKPOPKRSLKRADTEE-----ELAFKQTPSAG---KAMHTK 1426
QY 1637 SLAGVSLMLR--LIMASLRMDMAKVPYPPGGSTRTETSEITETTELIRRDVGPYGR- 1694
Db 1427 PAVGEKDDITFLGTGYQKLDQPGNLP--GSNRRLOTRREKAQALEEL-----GFRE 1477
QY 1695 -----FEYCIRKIIC--PIGVP--EPKETEPTQKGLRSS-----ALPKR 1732
Db 1478 LFQTPCTDNEPTADEKTKTKILCKSPQSDPADPPTNTKQRPKRSRLKRDVDEEFLARKLT 1537
QY 1733 P-----EPKQPGVYIETVWAVEELELMEIIRAF-----AEVE-----K 1767
Db 1538 PSAGKAMHTPKAA-----VEEEK--DITFVGTPEYKIDLGNLPGSKRRROTJK 1585
QY 1768 EKAQAVEOQAKRLLEQOKPTVIATSTTSPTSSTSTISPAQKVMAPVIGSVTTGKRWL 1827
Db 1586 EKAQALEDLAGEFKELFOTPGHTEESMTDKITEVCSKSPQDPDVKNP-----TSSKQRL 1639
QY 1828 TTYVGSATVTFQOONKHFOTFATWYKQGSNGVYQVQOKVL--GIIPSTGTSQOTPT 1885
Db 1640 KISLQK-----GVKKEEVLVYKGLQTSQTKQOT-- 1668
QY 1886 SPQPRATVTRIRNTSGSGT-----TSNSQVI-----TGPOIRGMVYIRPT-- 1928
Db 1669 -----HNETADGSGIAFKESAKQMLDPANYGGMGRWP-----KRPKEAOS 1712
QY 1929 LQOSTLEKAIIRTPVWQVPAPOQVMTQIIRGQVSTAVSAPNTVSSTPGOKSLTASIT 1988
Db 1713 LEDLAGFKELFOTPDHTEESTDDKTKTI-----ACKSPPEESMDPTST 1757
QY 1989 SNIOSSASQPPRPQOQOVKLT--MAQLTOLTQGHGNGGLTVYIQOGGOTTQGLQILP-- 2044
Db 1758 -----RRRPKTPPLGRDIVEELSAKLQLTQ-----THHTDKVPGD 1792
QY 2045 --OGVTVLPGBOOLM-----QAAMPNGTVO-----RFLF--TPLATTATTA 2082
Db 1793 EDKGINVFRETAKOKLDPAASVYTGSKRQPTPRPKAKPLADLAGKLEFOTPVCTQKPT 1852
QY 2083 STTTTVSTTA-----AGTGE--QROSKLS--PQMYQHDQTLTPPAOSSSYGPAKAPQTA 2134
Db 1853 HEKTTK IACRSQPDVGTPTIEFKPOSRLKRDVDEEELALRRKTPSGKAMDTPKPA 1912
QY 2135 -----QPSARPOPOQOPQAPQEVQTOPEVQOTQTVSS-----HV 2170
Db 1913 GGOEDKMAKMGTPVQKLDLPGLNPGSKRMPOTPEKE-----KAQALEDLAIFKELFOT 1965
QY 2171 PSEAOPHTNAOSSKFOVAAQSQPOSNVQOSPVYVQSPQTRIRPS----- 2215
Db 1966 PGTDKPT--TDEKTKTIACKS--PO-----PDPVDPRPASTKQKPRKMLRKADVEEFLAL 2016
QY 2216 ---TPS-----QISPGQSOVQOTTTSQPIPIOPHTSL--QIP--SQGQPOSQP----- 2256
Db 2017 RKRTPSAGKAMDTPKPAVSEKININTEVTPVQKLDLGNLPGSKRQOPQTPKEKAQALEB 2076


```
;; TITLE OF INVENTION: Variants of alternative splicing
;; FILE REFERENCE: 129181.4 Compugen
;; CURRENT APPLICATION NUMBER: US/09/724,676
;; CURRENT FILING DATE: 2000-11-28
;; NUMBER OF SEQ ID NOS: 97222
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO: 69127
;; LENGTH: 4189
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-724-676-69127

Query Match      2.7%; Score 390.5; DB 5; Length 4189;
Best Local Similarity 16.7%; Pred. No. 0.011;
Matches 580; Conservative 474; Mismatches 1179; Indels 1241; Gaps 129;

QY 17 OOSEDEDEDEDDDDDDSYREMEDEDDDDASYCTESSRSSTSYSTGRKRRPRNHR 76
DB 1061 KETESDODEIKETDRROSF-----ASLALKRKYSLTEPGMEDEYRSNK 1106
QY 77 SP-----ILEEKDIPLE----- 89
DB 1107 IPRHLLIALLFYFTIPIVLNDSDOYSAMASQVRLHLKFLFYLAISFVKINMVCPT 1166
QY 90 -----FPKSEDLAVPNEHIM-NVIAI-----YEVLN 116
DB 1167 FSNQIHRRVNYIFSKSIDVASPIQILTDNVEADKCGVTISIQYGFWMYPGYSSSH 1226
QY 117 FGTVALRSPFREDPCALVSOEGCTLAEMHVLLKALREEDTSNTFG-----P 168
DB 1227 GSYALKRAGIOYVFLSNLS-----DYSRVSGEYINMTPP 1264
QY 169 ADLKSVNSTLFIIDQMTPEVLRVYCESDKENHVLRYQ-----EAEDRYGVEKIKY 224
DB 1265 ASPKSNIM-----YSSSLPKSIITSAAPLISPLKAVSP 1301
QY 225 LQFLYDQFLITIAEELMSEGYIQDHCRCVCHKIGL-----LCCTCSAVNH 275
DB 1302 YKSRVDVSSAKITMASSISPVKQMGHAEVALVNGSISPLKASSSTLINCCKATATL 1361
QY 276 ECVKRPLEEVPEDEMCEVCA-----HKVPVTDVCVAEIQKNKPIYINHEPGRYSRR 329
DB 1362 Q-----EKISSATNSVSSVSAADTYEKFVSTTDCNA-ITPTQYICCTSISEFSVSN 1414
QY 330 KY-----NFLN--RRLIEEDTENENENKKIWTYSTVQVLAELIDCLDYWEAELCK 379
DB 1415 SRRKCTLYIPVYNICNYLNCFINVNSARYVNVLPALPKLPDINSFTKSAALLS 1474
QY 380 ILEEMREEIH--RHMDITEDLTNKARGSKSFLA-----AANEELIESIRAKK 425
DB 1475 PKTLTTEHPQHPFSRSTSPV-----KSSFLAPSAKLSTPSSLSOSSOELLKDAEMK 1529
QY 426 GDIDVKS-----PEETEKDNENTENDSKOAEKNREEDDS--LEKD----- 466
DB 1530 EDLMRTALLOTDRBEKRPQELPRKEGRIDDEEPKITYEKKVEDLVKVELLKRCVVD 1589
QY 467 -----SDDK--TPDDPEQKSEVGFKESEKSNGLSESPG-----AGKAS-- 506
DB 1590 NKSPKSPKSDKGHPEDDWIEFSSP--EIRARQOAAASQSPSLPERQYAKAASEKD 1647
QY 507 -GSTIILIRLNP--DSKISQLK-----SQOV--AAAANEAHLKFEKGEVYL 548
DB 1648 YMLTIVILYLTNDIGSSSLTNLKYFEDAKKDEGQKRLPRAIALQEHKIMPRASMR 1707
QY 549 VVNSGGEISRLSTK--KEVIKKNINNYFKLGQEGKRYVYINQYSTNFAALNKDHRD 605
DB 1708 TSTSKEELCKMADSEFGDTILIES--PDQFSQHDQKSPLSDSGFEETRSEKTPSAQSAE 1765
QY 606 HDRRHLLAHKECLTPAGEEKWANGSVHSGVLTISTRLTLTLENNI--PS-SFL-- 658
DB 1766 TTGPKRLHEVNPVPIPTETREVVHVIRSYDPSAGDVQIQTPREEVSKPSPFTHELEP 1825
QY 659 -PNWASHRANMIKAVOMCKSPREFALALALECAVAVPVVMPDIRWEFLGHTRLHMTSIE 717
```

```
DB 1826 KPTTSIKERF-KAFQ----- 1846
QY 718 REEKEVKKKKKKOEDEETMQATWVKYTERPVYKQWQKGEERYVTGYGMSWIKTHV 777
DB 1847 EDDHNVLSKGRVKEETHITTTTRAVYHS----- 1877
QY 778 YRFVPLPGNTVNVYKRSLEGRKNMDEMDSKRCKSRPK-----KIKIE 825
DB 1878 -----PG-----EGASERIEETMSVYDIKAKAFQSGNDPSKLAGLFEHKSAVS 1921
QY 826 PD-----SEKD-----EV--KSDA-----AKGADQNMEDT- 849
DB 1922 PVDHKSAAETSAQHAENKQMKRLEIIEVHLEKNOAEPVEVILIRETKHKEEMYYV 1981
QY 850 SKTTEKDDVVELLDSDSK--PCKEERPEVDDMT--ESHNCOESSOVVVNVEG 906
DB 1982 QKDLSGDITNLKDFLEPKHDAPPCSEGOQOEELTAEBSLPSTYLESSRVN--TPVSQEE 2040
QY 907 HLRTSYK-----KTKSKL-----DGLERRIQFTLEKORL-EKIKL----- 945
DB 2041 DSRPSSAQLISDSYKTLKLSHSTIEYHDELSELRGESYRAEKMLSEKLDVSHDT 2100
QY 946 -----EGIGIGKSTNSSKNLSESPVITKAEBCGOSDSMRQOSPANNQDP 994
DB 2101 EESVTDHAGPSSSELOGSKDRSREKATAPKKEILSKIKYDVSENGVGR-----VSKDEH 2155
QY 995 EDLQCSQSDSVLMSDPST-----TNKLYPKDRVLDVYSIRSPETKCP 1041
DB 2156 FDKVTVLHYSN-----VSSKHAMMRFTEDRLDRGREKILYIEDRV--DPTVAEAEKLT 2209
QY 1042 KONSIEDIEEKVSDIASGOEP-TSKTKGNDFFIDSKLASADN--IGTLICKNKK- 1096
DB 2210 EYSQFPRDKTEKND--ELQSPKKAHPKNGKEYSSQPTSSPKVLTTELLANDEN 2266
QY 1097 -----PLIOESDITVSSKSALHSSVYKSTNDRDAT-----PLSRAMD- 1135
DB 2267 VKARQHPGQGFPAKEEKAPSLPSSPEKRVLS--QOTEDSKSTVEAKGISISQKAPDG 2323
QY 1136 -----FEKGLGQSESNSTLENSDPVS-----IDDS----- 1162
DB 2324 POSGFOLKSKSLSTLKEQGTNAKSKDMSEDRKSDQSRIPVAKKIOESKLPYQVFA 2383
QY 1163 -SEEDMIYONSNEISEQ-----FTRQEDVEYLE----- 1191
DB 2384 REKQAKAIDLPOESVSVQDFWLKTKDEHAQSNELVANDSGDNVKKQRTKEMSKAMP 2443
QY 1192 -----PLKCELVSGESTGNCEDRLPYKGTBANG--KKPSQAKLLEPRVYKCSDOJK 1241
DB 2444 SFSEDOAKDLACHITSDLATRGPMDKKVFRTWESSGATNNKSQEKLSHVLVH----- 2496
QY 1242 LKMTTKKNNE--NRESEKKGORTSFFQINGKDNKPKIYLAGCELCIEISRVVSGNVEP 1299
DB 2497 -----DVRENHIGHPESKVDQNMENMSTYERERK--LITNGSLSEIKEMTVKSPS--- 2545
QY 1300 KVNINIKIIPENDIKSLTYKESAIRPFINGDVIMEDFERNSSSETSH-----TL 1349
DB 2546 -----KKVLIREYVVKGDHPGGLDDPSRNRSSASAASHIPVRADERRL 2591
QY 1350 SSSDAEGNVDLSLETLPSRKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKYT-- 1406
DB 2592 SSNIIPDGCEQS--APRKHELSQLQSSMSKMETVETQHFNSIEDEKVVYSEISKVSKHQ 2649
QY 1407 -----SPTITSEE-----ESNLSNDEIDENGPIPNKNEV-----NGES 1440
DB 2650 SYVGLCPLEETETSPKSPDLSFSPGKESPSVDVDSH--PIDLEKLAPLALQTEGSK 2707
QY 1441 KRRTVTEVTMTSTVATSEKTVIKVEKGDQTVVASTENKASKYVTTTTVTYTKLSTPS 1500
DB 2708 EIKTLPLVYV-----SFVQVQKQYEKEIQGQGVKKIISQDEKTYOETGTGTYTRQOQOPS 2763
QY 1501 TGSVDIISVKQSKTVTVTVTYVDTLTTTGGLVTSMT-----VSKYSTRDVKVLMKF- 1554
```



```
Db 2764 PQSGEDDTLEQVS-----FLDSSGSKPLPTEPSESEVSEYFTSKTPDLSLAVI 2813
QY 1555 -----SRPKTRSGTALDPSYRKFYTKSKSTLV--LPNDLKLARKGCI 1598
Db 2814 PGKPSPIPEVSESEEEQAKSTSL-----KQTVVEETAVEREMPNDSKSDNORPKN 2866
QY 1599 REVYFNYNNAKPLD-----KSLAGVSLMLRLMAS-- 1651
Db 2867 NRYATIEFPPLDADQIESDKHHYLPKEKVDMEVNIQDEHDKQLAEPIYRQP-P 2925
QY 1619 SP-----RPTFGITWRVRLQTV-----KSLAGVSLMLRLMAS-- 1651
Db 2926 SPVPBGADVSDSDDESITYQVPVVKYTFKLKEVDDEQEKPKASAEKASNOKELESNGS 2985
QY 1652 -----LRDDMAKAVPRGGGSTRTEFTETLTETTEIIRRD-----VGPGRIFE- 1696
Db 2986 GKDNREGLGIDSPQNEIANGNDOSITECSIATTAEFSHDTPATEIDSDGDLDEDED 3045
QY 1697 -----YCIKRIIICPIG-----VPETPEKPTPQPKGLRSS 1726
Db 3046 GLTESDSKLPQAMEIKKDDIWNTEGLIKRADBSFSOSKLEVIEBEGVGDDEDKPSPKSS 3105
QY 1727 ALRPKRPEPKQTGVIIETVAEELEWEIRAFAREKEKAQAVEQA----- 1777
Db 3106 SSEKTPDKTDQKSGAOFETLEGHNPDRSVFPTDYFSYKVDDEFATPEFKTATGLDPEDPW 3165
QY 1778 -----KRLQOQKPTVIATSTSTSTSTSTSPACKVMVAPISGSVTTGTM 1825
Db 3166 SNNRGDDEVFDSKREDEKPRGLAVEEDRSP-ATTPTTPPAR-----TPT 3209
QY 1826 VLTTKVSPATVTFQOQKNFHQTFATV-----XOGQS 1858
Db 3210 DESPTPEPRPFHEKMEEMTRSGAIDMSKRDVEERLQFQIGHTEGKSGDQGE 3269
QY 1859 NSGVQVQOAVLGLIISSTISQOTFSPQRAIYVIRN-----TSGSGGT 1906
Db 3270 DKSMVATPQ-----POSGTVE--TNLERNVETPVEPNPSIPISGECQETSSGSL 3322
QY 1907 TNSOVITGQIRPGMIVITPQOSTLGAIRTPVMVOPGAPQOVMTOI-----IRG 1960
Db 3323 EKSAATNTSKVDP--KLTPPIKM-----GISASTMTMKKEGGETLDEIVAMVSCOG 3374
QY 1961 QPVSTAVSADNTVSTPG-----QKSLTSAT-STNSIQS-----SAS 1996
Db 3375 LENETITMISNTANSQMGVAPRHEKHDFQKDNFNNNNNLDSSTIOTDNIMSNIVLTHEAP 3434
QY 1997 QPRPQOGQVKLTMAQTLQTLQGHG-----GNQGLTVVIOGQCG----- 2035
Db 3435 TCTTEKDNPVKSSGKKTGVLOGHCVARDKOKVLGEQOKTELIGIRKSKPLIKATISPKD 3494
QY 2036 -----TTGQLQILPQGVVILPGPGQQLMQA----- 2060
Db 3495 TTFPNHMSNTKASKMKQVSOSEKTKALTITSSCDVKSRIIVKMTPRNIIIAVKKACATOK 3554
QY 2061 -----AMPNGVQRFLETPPLATTATTAFTASTTTTTVSTTAAG----- 2095
Db 3555 QGQPEKAKOLPSKLPFVKVRSCTVTTTTTATTTTTTTTTTTTTTCTVVKRSOLKEVCK 3614
QY 2096 -----TGE-----OROSKLPQMOVHODKTLF-----PA----- 2119
Db 3615 HSTIEYFGISGETLKLVDRLSEEEKKMOSELSDEEEKYLLKHVVYVRFPGMPAFGYKKA 3674
QY 2120 --OSSSVGPAKAPQTAQPSARPOQTAQPSAPQAEVQTAQPEVQTAQTVSSSHVSEAOPT 2177
Db 3675 RDKTEAPLPIKSKSEKSGSRRTGPGSPCE-----RDIIMAIAD----- 3719
QY 2178 HAOSKQVAAQ-----SOPQS----- 2194
Db 3720 HLGISMTLARELNFVDEIINOI RVENPNSLISOSFMLKRWYTRDKNATTDALTSVLT 3779
QY 2195 -----NWOG-----OSPVVQSPQSORIRPST 2216
Db 3780 KINRDIIVTLLEGPIFYGNISGTRSPADENNWFHDPVDGYPQLQVLELPTGLHTYPTPT 3839
```

```
QY 2217 PSOLSPGQSQVQVQTTISQPIPIOPHTSLQ---IPSGQ-----QPSQPOVOSSNQTLSSGQ 2269
Db 3840 PFO-----QDDYESDISIESPTEPSRLSDGLVPSOGONIEHSDAGPVPYTAERASLEDK 3895
QY 2270 TLNQVSVSPSPRQOLIOQOPQVIAVPOLOQOVVLSQIOSQVVAQIAQOOSGVPOQIK 2329
Db 3896 LEDSVPLTE-----MPEAVDVDEQLENVCLSE-----YPQYLGNIAGSPKDYK 3939
QY 2330 ----LQPIQIOQSSAVQTHQIONVTVQAASVQEOLOQVQOLRDQOQKKQOQIEIKRE 2385
Db 3940 PAEPKRLGVSSQEOEKSGSPDEEMBEKLSLFEQIOLEEGVSEBMEEEKVQALIKR- 3998
QY 2386 HTLQASNOSEIIOKVYMKINAV-----IEHLKQKSMTPA--BRENEMAIVCNQVWKYI 2439
Db 3999 -----VQOAELEMSSTITGMONETSSGNLESCAQARVYTGGLDRDDSPQCGDSTISYL 4053
QY 2440 LDKIDREE-----KQAKKRKREESVEQ--KSKQNAVTKLSALLFKHKEQLRAEIL 2488
Db 4054 KGEAGKFEANGSHTEITPEAKTKSYFPESQNDVGKQSTKETLKPILHSGHVEEPASPLA 4113
QY 2489 KRALLDKQDQIEVQELKDKLIKKEKDLMOQAQATVAVAPCPVTPVLPAPRAPPSP 2548
Db 4114 AYQKSLSETSKLITIE-----TKPCVPVSMKMKMSTSPADGK 4150
QY 2549 PPPPGVQHTGLSTPLPLVPAQKRKRREKSSSKSKKMMISTSKETKKDKR 2602
Db 4151 P-----RLSLHEBEGSSGSEBQKQSGCFV--KTKKEIR 4181

RESULT 7
US-09-724-676-69160
; Sequence 69160, Application us/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69160
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-724-676-69160

Query Match          2.7%; Score 388.5; DB 5; Length 3259;
Best Local Similarity 16.8%; Pred. No. 0.0096;
Matches 576; Conservative 493; Mismatches 1207; Indels 1155; Gaps 129;
```


Db 415 YASSSTLINGCATATLO-----EKISATNSVSVSAATDVEKYSTTDCA--ITP 467
Qy 313 NKPIYRHEPIGYDRSRKY-----MFLN--RRLIEEDTENENKKIMYSTVOA 362
Db 468 TQVIGECSTISFVSNSFRKCTLYIPWVVICNYLICNFINVNSAISVNVNLPALP 527
Qy 363 ELIDCLDKYWAELCKLIEEMREEIH--RHMDITEDLTNKAAGSKSLA----- 411
Db 528 KLPDSNSFTKSAALSLPKTLTLETHERPOHFSRTSPV-----KSLFLASALKSLP 582
Qy 412 ---AANEETLESIRAKKGDIDNVKS-----PEMEKDNENENDSKAEKR 455
Db 583 SLSOSSQELTKVAAEKEDMPTALOTDVEBEKPOPELPEKGIIDDEBEPIYEEKV 642
Qy 456 BEFEEDQS--LEKD-----SDK--TPDDPEEGKSEVGFKSEKNGELSEP 499
Db 643 EDLVVSEILKDYCDVKNKGSPKSPKDGHPEDDIEFSS--EIREARQOAAASQSP 700
Qy 500 G-----AGKAS-----GSTRIITRLNRP--DSKLSQK-----SOQV--A 531
Db 701 SUPERVOYAKAKASEKDYMLTKYVIDYLTNDIGSSSLTNLKYFEDAKKDGEGOKFVLRP 760
Qy 532 AAANEANKLFKFGKEVLYVNSOGELSLSTK---KEVIKGNINNYFKLGEGKYRVYHN 588
Db 761 AIALOEHKLMPPASAKRTSTSEKELCKMADSPFGDTILES--PDQSOHODKSPLOS 818
Qy 589 QYSTNSFALNKHQHRHDHDKRRLAHKFCITPAGEFKMNGSVKSLTISTRLITOL 648
Db 819 GGETNSEKTPSAPOSAETTGPKPLPHEVP IPRVITETREVVHVIRSPASGDVQTOP 878
Qy 649 ENNT--PS-SFLH---PNMASHRAMIKAVOMCSKPREFALALALECAKPVVMDPI 700
Db 879 BEPVSKRSPPTMELEPKPTTSSIKEF--KAQOM----- 911
Qy 701 WREPLGHTLHMTSIEREKEKVKKKKEKKEBEETMOATWYKTFPVKHOVWKQGE 760
Db 912 -----KASEEDHNRVLSKGMKVEETHITTTTRVYSP----- 947
Qy 761 YEVTOGWSMTSKTHVRFVPLPNTNVNRYRKSLEGTAKNNMDENDESDKRCSPK 820
Db 948 -----PG-----EGASREETMSVHDIMKAKQSGR 974
Qy 821 -----KIKIEP-----SEKD-----EV---KGSDA--- 838
Db 975 DPKELAGLFEHKSAVSPVHKSAETSQAQHAKEKQMKPKLERILEVHEIKQNAEPT 1034
Qy 839 -----AKGADQEMDI-SKITEKKQDVKEILDSDK--PCKEERPEVDDDKMT--ESHV 889
Db 1035 VIIRETKHKPEKEMVYQKDLRGDINLKDPLPEKHDAFPCCSEOGQOEELTAEESLP 1094
Qy 890 NCOESSQVDVNVVSGFHLRTSYKK-----KTKSSKL-----DGLLEIRIKOFTL 934
Db 1095 SYLESRRVN--TPVSOEEDRSPSAOLISDDSYKTLKLSOHSTIEYHDDSELRGSYNF 1153
Qy 935 EKKOFL-EKIKL-----EGIKGIGIKTSTNSKNLSSESPVITKAREGO 977
Db 1154 AKKMLSEKLDVSHSDTEESVTDHAGRPSELOGSDKRSREKILATAPKKEILSKIKDVS 1213
Qy 978 SPSMBOEQSPNANNDOPELLIGCSOSDSSVLRMSDPST-----TNKLYK 1024
Db 1214 ENGVOG-----VSKDEHFPKAVVLAHSGN---VSSPKHAMMMRTEDRLDREGRELLIYE 1264
Qy 1025 DRVLDDVSIIRSPETKCPKONSIENDIEEKVSLDASGOEP--TKSKTKGNDFTIDSKLAS 1083
Db 1265 DRV--DRVTKBAEKLTEVSOEFFROKTEKLN---ELOSPKKARKPKNKEVSSQSPSS 1319
Qy 1084 ADD---IGTLICKNKK-----PLIOESDTIVSSKSAHLSVPKSTNDRD 1126
Db 1320 SEKVLTELLELANSDEWYKAROHGPDGQGFPAKEEKAPLPSPEKMYLS---OQTEDESK 1376
Qy 1127 AT-----PLSRAD-----PEGLGDDSESNLLENSDDVVS-- 1158
Db 1377 STVEAKGSIQSOKAPDPOSGFOLKOSKLSIRLKFEOGTHAKSKDMSOEDRKSQOSKRI 1436

Qy 1159 ----IODS-----SEEDMIVONNESISEQ-----FTRDEOVEYLE----- 1191
Db 1437 PVKIQESKLPYQVAREKQAKIDLPEDESYSVQKDFMLTKTKDEHAOSNEIYVANDSGS 1496
Qy 1192 -----PLKEIVSGESTGCEDERLPVKGYEANG--KKPSQ 1224
Db 1497 DNYKORTEMSSKAMPDSESEOGAKDLACHITSDLTARGMDKVRFTWESSGATNNKQ 1556
Qy 1225 OKKLEERPYNKCSQDQIKLKNITDKNNNE--NRESEKGOFTSTFOJNGDKNKRIYKGE 1282
Db 1557 KEKLSHVLVH-----DVRENHIGHPESKSVDOKNEMASTERERK---LLTNG 1601
Qy 1283 CLKIEISERVYSGNVBPKNVNNIKIIPENDIKSLTVKESAIRPPIGVDYIMEDFNERNS 1342
Db 1602 SLEIKEMVTKSPS-----KVLVREYVYKGDHPGGLDDPSRRSES 1644
Qy 1343 ETKSH-----LLSSDABGNRDSLETLPSTRKESDSTOTTPSPASCPSNSVNOV 1392
Db 1645 SAVSHIPRVADERRMLSNIPDGFCEQS--APPKHELQKLSQSSMSKETVETOHFNSI 1702
Qy 1393 EDMIEITSEYKVT-----SPITSEE-----ESNLSNDFIDENGLP 1429
Db 1703 EDEKVTYSELSKYSKQSYVGLCPPLLETETSPTKSPDSLFEPSGKESPSDVFDHS--P 1760
Qy 1430 INKNENV-----NGESKRRTVITEVTTMTSTVATESKTIYKVEKDKOTVSVSTENCAR 1483
Db 1761 IDLEKLAFLAQEGGEKELTLPVYV---SFVOVGQYKEKEIQCGVKKIISOEKTVO 1816
Qy 1484 STYTTTTVTYKLTSTSTGSDVUIISVKEOSKTVVTTVYDLSLTGCGILVYSMT----- 1538
Db 1817 ETRGFTYTPROOKOPSPGSPGSPEDDYLEOVS-----FLDSSGKSPLETPETPSSEE 1866
Qy 1539 VSKEYSTRODKVKLMF-----SRPKTRSGALPSYKFPVTKSKKSLFV- 1583
Db 1867 VSTEFTSKTPDSLIAYIPGKRSPPIPEVSESESEDOAKTISL-----KQTVBEETAYE 1919
Qy 1584 --LPNDLKLARKGIREVPYENYNAKPALD----- 1613
Db 1920 REMPNVDSKSNQRPANNRAYIEFPPLDLADQIESDKHHYLPKEKEDMLEVMDQE 1979
Qy 1614 -----IWPYSP-----RPFGITWRIRLOTV-----K 1636
Db 1980 HDKYQALAEPIYRQOP--PSPVPGADVSDSDDESIIQOPVPVKKYTPFKLEVEDEOKEKRP 2038
Qy 1637 SLAGVSLMLRLMAS-----LAMDMAKAVPPGGGSTRSETETITTEIILKRRD 1687
Db 2039 ASAEKASNOKELESNGSGKDNFGLGLDSPONETIAONGNNDOSITRCSIAATTAEFSHDT 2098
Qy 1688 -----VGPYGRFE-----YCIKILICPIG-----VP 1709
Db 2099 ATFIDSLDGDIDDEDGILTESDKLPIDAMEIKKIDWNTREGILKAKADRFSQSKLEVIE 2158
Qy 1710 ETPKETPTPOKGLRNSALRPKRPEPTKOTGPVILIEWAAEELMEIRAEFERYEKEK 1769
Db 2159 BEEGKVPDEKPKPSKSSSEKPTDKDOKSGAOFTLLEGHNPDRSVFPPTYEFYKVDIEE 2218
Qy 1770 AQAVEGOA-----KKRLEQOKPVIANTSTSPSSSTISPAQ 1808
Db 2219 ATPFKTVATKGLDFDWSNNRGDEVFDSKREDETRPGFLAEDKSP--ATTPDTTPAR 2276
Qy 1809 KVVAVPISGSVTTGTMKVLTTKVGSPPATVTFQONKNFHOTFATV----- 1853
Db 2277 -----TPTDSTPTSEPNPFPFHGKMEFMRGSAIDMSKNDVFEERLOFF 2322
Qy 1854 -----KOGOSNGVVOVQOKVLGILPSSGTSGOQFTTSQOPRTATVITPN-- 1899
Db 2323 OIGEHNTSEKSGDGGGDSMTATPQO-----POSQDTIVE--TNLERVNETPTEVPNS 2375
Qy 1900 -----TSGSGTTSNSOVITGOAIRPGMTVTRTPOLOSTLKAIRFPVWVQOPA 1949
Db 2376 IPTSGEOGEGTSSGSLKESAAATNTSKYDP---KLRTPIKM-----GISASITMYKKKEG 2427

QY 1039 -----KCPKNSIENDIEEKVSDLASGQEPYKSKTGNDFIDDK-LASADD 1086
Db 935 EKDTPELMERLEVTREKLE--LSQRLSDLS--EOLKQKPEISFLMEYEVSKLPEKE 988
QY 1087 ICTLICKNNKPLIOESDPIVSSKSALHSSVKTSTNDAPPLSAMDFECKLGDSDS 1146
Db 989 PVSILRR-----ELETIINHNR-----ENVQSCDTQV 1016
QY 1147 NSTLENSDPTVSIODSEEDMIVONSESI SPOFTRBODVEYLEPKCELYSGESTGNC 1206
Db 1017 SSLT-----DGVYMTSRGAGSVSKVKSFGESKIMVEDKYSFE--NMVGEESKQE 1068
QY 1207 E---DRLP--VGTENANGKPPSQOKLEBRPVNKCSDQIKLNTDKNNENRESKQ 1261
Db 1069 QILBHLPSVTKESSLRAQOPSENDLQ-----KELNVLSKSEONDILR-LQMEAQ 1116
QY 1262 R-----TSTFOI-----NGKDNPKIYLKGECLKEISESVVSGNTEPKNNINKII 1308
Db 1117 RCLSLVYTHVDQVREYENEDK-----KALCSLKEELIFA--QEEKIKELQK-I 1164
QY 1309 PENDIKSLTVKESALRPTNGD-----VIMEDFNERNSEETKSHLSSDAEGNYRDL 1362
Db 1165 HOLEIQTMTQET-----GEGKPLHLLIGLKQKAVSEECYFIQTLCSVLGEY---- 1213
QY 1363 ETLPSKESDSTQTTTPSASCPEBSNVNOEDMEITSEVKVYTSPTISEESNLNDF 1422
Db 1214 -----YPLALKC-----EVNAEDKEN-SCGY 1233
QY 1423 IDENGLPINK--EAVNGESKRTVITEVT-----PMTSVNAESKVIIVEKG 1469
Db 1234 ISENEPELOADRYEVOQDQENHNTLKVTEBYNKLVLQFLSKIMQOQDGMKLEFG 1293
QY 1470 DKQTVVSTENCASKSTVTTTTVTVTKLSTPGSVDIISVKEOSKTVVTVTVDLSTT 1529
Db 1294 E-----ENLPKE-----ETEFLSHSQ----- 1310
QY 1530 GGTLVTSMTVSEYSTRDKVYKLMKSRPKKTSNGALPSYRKFFVTKSTYKSIFVLPNDI 1589
Db 1311 -----MTNLBIDIVNHSKSLSDLODEKTKLEQVOELSLISSLOOQ-----L 1354
QY 1590 KKLARKGIREVYFMYNKKPALDIMPPSPRTFGITWRYLQVYKSLAGVSLMLRLM 1649
Db 1355 KETEO-----NYEAE-----IHCLQ--KRLQAVS----- 1376
QY 1650 ASLRMDMAKVPYGGGSTRTESETETITTELIRKRDVPGYIREYCIRIKTICIGVP 1709
Db 1377 -----ESTVBP-----SLPVDSVITSDAQRTMYPGS-----CVKKNI--DGTI 1414
QY 1710 ETPKETPTQKRGKSSALRPKRPEPKOTGPYIETWYA--EBELELMEIRAFEREVERE 1768
Db 1415 EFSGEGVKEETNIYKLEKQYQOELEEVAVYIVMSIAFAQOTELSRISGKENTASS 1474
QY 1769 K-AQAVEQOAKRLEBOKPTVATSTSTSTSTISPAQVMAVAPISGSVYTGKMWL 1827
Db 1475 KQAHAVCOQEOHYFNBK-----LSODQIGFOFETVDVKFEERKPLKLEGEHCKEL 1529
QY 1828 TTKVSPATVTFQONKNFQTEFATWVKQOSNSGVVOQKVLGILPSSTGTSOQFTSF 1887
Db 1530 LS-----NSDPHD-----IPESKDCLITI-----SEKHFMSD 1556
QY 1888 QPRTATVTIRPMTSGSGGTTSSQVYITGPQIRPMTVITPILQOSTLGAIIITPVMVOP 1947
Db 1557 KPIFVROSITHDEISVSMASQMLNEBOLDMROELROYOEHOQATELLROAHMRM 1616
QY 1948 GAPQOVMQIINGOPVSTAVSAPNIVSSPGQKSLTATSTNIGSSASQPRPQOGVYK 2007
Db 1617 EKQREDOEL--QEBIKRLNQLAORSSIDENELVSEERVLLELELKLKOLSLAGREK 1673
QY 2008 LTMALUTLOLQHGNGNOGLTVYIOGGQTTGQ-----LQILPQGV--TVLPQPGQOLMQ 2059
Db 1674 LCCELRNSSTQONGN-----ENQGEVEBQTFKEKELDRKEDVPPRILSNERRALQK 1726
QY 2060 AAMPNGIVORFLFTPLATTATTAST-----TTTVSTTAAGTGEOGRSKLSFOMOV 2110

Db 1727 A---NNRLKILLLEVAKTTAAVEETIGRHVGLILDRSSKQSSASLIRSEABASVSCV 1783
QY 2111 HODKTLP-----PAOSSVGP-----AKAQOTQOPSAQPQOPQOPAPQE----- 2153
Db 1784 HEHTRVDESIPSTYSGSDMPFRNDIMMSKVTEEGTGLSORLYVRSGFAGTIDPEMBELM 1843
QY 2154 --VQTOPEVOTQTTVSSHVPSEAPTHAOSKQVQAOS-----QPOSNVQOSPVYVOS 2206
Db 1844 LMSIRLQAVERKLEBIAISPTSSQLEHAKVQTELRESRQKQKQATESLKQCEBLERRL 1903
QY 2207 PSOTRIRPSTPSQLSPQ-----OSQVOTTT-----SQPIPIOPHTSLQIP 2247
Db 1904 HESRRAREQVLAVLSKAEVIGDYADEKTLFEFOIQIEKTDIIDRLQELLCASNRLOELE 1963
QY 2248 SOQOPQOPQOVSSQTLSSGOTFLNQVSSPSRPQLOIQOPQVYIAPVQLOQOVVLS 2307
Db 1964 AE-----QOQIOERELSLR--QKAMKAEAGP--VEQ-----QLQJTEKLM 2002
QY 2308 QLOSQVVAQIOAQOSGVPOQIK-LQPLPIQOQSSAVQTHQIQNV-----YTVQASVQEO 2362
Db 2003 KEKLEVQOQAEKVRDLOQVAKLEIDVEQVSRFIELEBCKETELMDLQOQALQOL 2062
QY 2363 QRYVQLRDQOQKKQOQOIEIKREHTLQASQSEI QK-QVMAKNA-----VIEHLKQK 2416
Db 2063 EKMRKFLDEQALDREHERVDFQOEIQKLEQOLKVVPFRFQIPISHQTREREYQOLANHLKEK 2122
QY 2417 ---SMTPAERENQRM--VCNQVMKYILDKIDKEEQOAKKRRESVEQK----- 2463
Db 2123 DKCSHELLSBOLOQDIOERNBETIEKLEFRVRELEBOLLEDRKNGAVKAPPELSLEVOL 2182
QY 2464 -----RSQONATKLSALLFKHKEOLRAELIKKRALDK-DLQJTEVQ-----EEL 2506
Db 2183 QAERDALDKREKETITULEQL-----EQPRELEENKMBEYQOLMQOIEQKKESTRLQEL 2238
QY 2507 KRDLKIKKEKDLMOQAATAVAAACPPVTPVLPAPAPPPSPPPPGVOHTGLSTPTLP 2566
Db 2239 EOENKLFKD-DMEKLGIA-----IKESDPMSTODOH 2268
QY 2567 VASQK-----RKREEKD-----SSSKSKKKMISTSKETKDKTKLJCQIKPYDSKPY 2617
Db 2269 VLFKFAQOIIQEKVEIDQLEQVTKLQOOLKITTNKVIYEENELIRDET-----Q 2321
QY 2618 IGCRCQNMVHGRVGILOSEAELEIDVYVQPCOSTEDAMTVLTPLEKDYEGLAKEV--L 2675
Db 2322 IEC-----LMSDQECVK-----RNREBEIEQLENEVIEIKLQOELANIGQK 2360
QY 2676 RSLQAKKMAAPLEVPDNDADPDYIGVIEKPMDLATMEERVOYRYIEKLTIFYADMTKIF 2735
Db 2361 TSMNAHSLS---EFADS-----LKHOLDVVAIEKLALEQOVETANEEMTFMKNVL 2407
QY 2736 DMCRRYVNSPSPPYQCAEVLESFFVOKLKGFKASRSHNNKLOS 2778
Db 2408 KETNF-----KMNQLOTELFLSKRERSVEKIOS 2436

RESULT 9
US-09-724-676-68917
; Sequence 68917, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/774, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68917
; LENGTH: 3910
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-68917

Query Match 2.7%; Score 385; DB 5; Length 3910;
 Best Local Similarity 17.2%; Pred. No. 0.013;
 Matches 481; Conservative 478; Mismatches 1036; Indels 808; Gaps 114;

QY 305 DCAVEI-----QKNP--YIRHEPIGYDSRRKYM-----FLNRLLIE 341
 DB 125 DCSEVNGCSFVMGTGKPTMLRREEFGVDYSSEGAODSPHLEMESELAGQHEIE 184
 QY 342 EDENENKRIWYXSTKYVLAELIDCL-----DKDYWEALCITL 381
 DB 185 ELNLEEMEMVTYOTBELOQLOFEFAIKORDGIITOLTNLQARREKQETMEFLELT 244
 QY 382 EEMRE-EIH-RHMDITDITLNKARGSKSFLAANEI-----LESIRAKG 426
 DB 245 EOSOKLOIOFOOLOASETLNSTHSTSAADLLQAKOOILTHQOOLEDHDLLEEDYOKKE 304
 QY 427 DI-----DNKSPETKDKNETENDSKDAEKNE-----FEEDOSTLEKDS 467
 DB 305 DFTMOISFLQEKIR-VYEMQDK-KVENSNEKEIOEKETIIEELNTKIIEEKKTLLEKD 362
 QY 468 DDKTPD-----DDPEQKSEVGFSEKSNGLSESPAG-KGASGSTRITIRLNP 518
 DB 363 KLTTADKLLELOEQIYQKNOELKNMTELITNSKOKEROSSEELKOLMGVIELOKRNHK 422
 QY 519 DSKL-----SOLKSOVAAAHEA-NKLFKEGKVLVYNSOGELSRISTKKEVI 566
 DB 423 DSOFEITDIORMEOETORKELOLFAELDEMYGOOIVOMKELLIRQHAQOEMKTRHKE 482
 QY 567 MKGINNYFKLOGGKRYVHNQYSTNSFALNK--HQRHEDHOKRRHLAKFCITPAGFE 624
 DB 483 MENALRSYSNI-----TVNEDQIKLMVAINELNITKLODTSOKEKLEL 528
 QY 625 KMGSVHSGKVLITSLRLTITOLENNIPSSFLPHNASHRAMIKAVOMCSKREFALA 684
 DB 529 ---GLIEEKCALOQLOEDLVEEL-----SEFREQ--IQARQTIAGE-SKINEAHKS 576
 QY 685 LAILECAVAVVMLPTREFLGHTRHMTS---IREKEKVKKKKEQOEBEETMOQA 740
 DB 577 LSTVEDIKAVIVSASESRKELEKHEAVENTYKIKLEMEKEKAVYDRAAESOEALER 636
 QY 741 TWKYTPPVKHOVKGE---EYRVYGYGWSWISTHYRRFPKPGMNVYRSLE 797
 DB 637 LRQOLFSHEBELSKLEDEIEHRIN-----TEKMDLGHYKOOID 680
 QY 798 GTRNNNDEND--ESDKRKSRSPPKIKIPDESEKDEVKSDAKGAD-----ONEM 847
 DB 681 GLQNEHMQKLETMOFEKDNLITKONQILFISKLKLOQSLVNSKSEMTLOINELQKEI 740
 QY 848 DISKITEKK---DOYKEL-----LDSOSKPCKEPEMEV 879
 DB 741 EILROEKEKGTLEQEOVLOLKTLELEKOMEKENDLOERFAOLEAE-NSILKDEKCTL 799
 QY 880 DDDMKTSHVNOCESSQVDVAVNVEGFHLRTSYKKKTSSKLDGLLERRIKOFTLEKOR 939
 DB 800 EDMIKITIPYSQEE-----RLIFLDSIKSKSKSVWEKEL-ELLIEBENED 843
 QY 940 LEK--IKLEGGIGIGCTSNSSKNL-----SESPIITAKAREGOSDSMROE---Q 985
 DB 844 LKOOCIOLENEIEKORNTFSFAEKNEFVNYQLOEYACILKVKDLEDLSKNQOLEYKS 903
 QY 986 SPNANNQPPDLIOGCSQSSSVLRMSDSPHTNKLKPKRYVLDVDSIRSEP----- 1038
 DB 904 KLAALNEE-----LHLORIIMPYTKM--KSSVDEDEKTFVAELMEQEV 946
 QY 1039 -----KCPKONSINDIEEKVSDLASGOEPTKSKTGNDFFIDSK-LASADD 1086
 DB 947 EKOTTELMELVTKRKLE--LSQRLSDLS---EOLKOKPGEISFLNEVKSLEKEPE 1000
 QY 1087 IGLICKNNKKPLJOESDITVSSSKSALHSSVSKSTWDRDATPLSRAMDEGLKCDSES 1146
 DB 1001 PVSJLRG-----ELEIILHNRA-----ENVOSCDTQV 1028
 QY 1147 NSTLENSDTVSIQDSSSEDMIVONSNEISIEQFRTREDDVEVLEPLKCELVGSGESTGNC 1206

DB 1029 SSLL-----DGVVMTSGAGGSVSKVKNKSFGEESKIMVEDKVSPE-----NMTVGEESKOE 1080
 QY 1207 E---DRLP--VKGEANGKPPSOOKKLEERPYNKCSQOIKLNTDCKNNENRESEKKG 1261
 DB 1081 QLIIDHLPVYKSESLSRATPSENDKLO-----RELNLKSEONDLR-LQWEAQ 1128
 QY 1262 R-----TSTFOI-----NGKDNKPKIYKGECLKEISERSVSGNPPKVNINIKII 1308
 DB 1129 RICSLVYSTHVDQVREYEMENED-----KALOSLKEBELLFA--QOEKIKELOK-I 1176
 QY 1309 PENDIKSLVYKESAIRPFINGD-----VIMEDFERNNSSETSHLSSSDABGNRDSL 1362
 DB 1177 HOELIOTMKTOET-----GDECKPLHLIIGKLOKAVSECSYFLOTLCISVGEY---- 1225
 QY 1363 ETLPSTKESDSTOTTPPSACSPESNSVNOYDEMELEFSEVKKATYSSPTIIEESNSLNDF 1422
 DB 1226 -----YTPALKC-----EVNAEDKEN-SGDY 1245
 QY 1423 IDENGLPINKN---ENVNGESKRTVITEVT-----TMTSTVATSKRYIKVEKG 1469
 DB 1246 ISENEDELQDYREYQDFENMHTLTKVTLEEYKNLLVLOTRLSKIMGOQDTGMKLEFG 1305
 QY 1470 DKQTVVSTENCAKSYVTTTTVTYTKLSTPSTGASVDIISYKEDSKTVVTTVTYDLSLTT 1329
 DB 1306 E-----ENLPKE-----ETFEELISHQ----- 1322
 QY 1530 GGLTVSTMTVSKSEYSTRDKVYKLMKFSRPKTRSGTALPSYRKPFYTKSKSIFVLPNDL 1589
 DB 1323 -----MTNLEDIDVHKKSLSLQDLEKTKLEBOVQOELSLISSIQO-----L 1366
 QY 1590 KTLARKGIREVPEYFNANAPALDIMEYPSRPFTGITMYRLOTYVKSLSAGVSLMLRLM 1649
 DB 1367 KETFO-----ANEAE-----IHCLQ--KRLQAVS----- 1388
 QY 1650 ASLRMDMAKVPVPGGSGSTETSETITTEITTEIKRNDVOPYIGIRFYCIKRIICPTGP 1709
 DB 1389 -----ESTVP-----SLPVSQVITTESDAQRTMPPGS---CYKKNI--DGTI 1426
 QY 1710 ETPKETPTPOKGRSSALRPKRPETKOTGPVIERVWA--EELELELFRARAVEREKE 1768
 DB 1427 EPSEFVYKKEETINIVKLEQYOQOLEEYAKYIVSISLFAOQTELSTRISGKENTASS 1486
 QY 1769 K-AQAVEQAAKRLQOKPVIATSTTSPTSTSTISPAQKVAVAPISGSVTTGTGMVL 1827
 DB 1487 KOAHAVQOQOEHFNEK-----LSQDQIGFOFFETVDVAFKEHFRPLSKELDEHGEKIL 1541
 QY 1828 TTKVGSPIATVTFQONKNFHOFPATVWAKQOSNSGVVOOVOKVIGIIPSSGTSGQFTFS 1887
 DB 1542 LS-----NSDPHD-----IPESKDCVLT-----SEMEFSKD 1568
 QY 1888 QPRTAVTIRPNTSGSGGTTSSNOVITGPQIRGMVYIRPPLQOSTLGKAIIRTPVAVOP 1947
 DB 1569 KTFIVKOSIHDEISVSSMDASROLMEBOLEDMRQELVAYQOEHQOATLLOAHNRQM 1628
 QY 1948 GAPQOVNTQIIRQOVPYSTAVSAPNTVYSTPGKSLSTATSTSNIQSSASQPPRPOQGVK 2007
 DB 1629 ERQREDEQOL--QOEIKRLNROLAQRSSIDNENIVSERBVLLELEALKQSLAGREK 1685
 QY 2008 LTMALQTLQLOHGNGGLTVIIOGQQTGO-----LOLIQGV--TYLPPRGOOLMG 2059
 DB 1686 LCCELNRNSTOTQNGN-----ENQGEVEEQTFKEKELDRKEDVPELLISNERVALOK 1738
 QY 2060 AAMPNGTVGRFLFTPLATATYAST-----TTVYTTAAGGEGOSKSLPOMOV 2110
 DB 1739 A---NNRLKILLEVKTITAAVEETIGRHVGLILDRSSKQSSASALIMREABASVSKCV 1795
 QY 2111 HODKTLF-----PAQSSVGP-----AKAPQOTAPSAARPQOTQOPAPQPE----- 2153
 DB 1796 HEHTRVYTDSEISPSYSGSDPRNDIMMKSIVTEGELELQRLVRSRGPAGTEIDPENBELM 1855
 QY 2154 --VOTOPEVOTQTTVSSHVPSEAOPTHQOSSKPOVAAQS-----QPOSNVQOGSPYRVRS 2206

Db 1856 LNISSRLAANVAKLLBAISSETSOLEHNAVOTELMRESFROKOEATEBSLKOEELREBL 1915
Qy 2207 PSOTRIREPSPOLSPGQ-----OSOVOTTT-----SOPPIOPHTSIQIP 2247
Db 1916 HRESNAREGLAVELASKEVIDGVADEKTLFPERQIOEKDIDIRLEDELLCSNRLQELE 1975
Qy 2248 SOGQPOPOVOSVOTLSSGOTLNOVSVSPSPOLQIQOPPOVYIAVPOLOQOVUS 2307
Db 1976 AE-----OOOIOEBELLSR-----OKEAMKAAGP-----VEQ-----QLLQETEKIM 2014
Qy 2308 QIOSOVVAVIOAOGSGVPOQIK-LQPIQIOGSSAVOTHOIONV-----VTQASVQOEDL 2362
Db 2015 KEKLEVOCAAEVRDQLOKALELIDVEOVSRTLEBQENKTELMIDLROQOALEKOL 2074
Qy 2363 QRYVOOLRDOQOKKQOQOIKREHRETLAOSGEIIOK-OVAKKHNA-----VIEHLKOKK 2416
Db 2075 EKMRKFLDEQALDREHERVPOQEOIKLEQOLKVPRPPISEHQREVEQOLANHLKEKT 2134
Qy 2417 ---SMTPAERENQRYT-VCONVMKYLIDKIDKEEQAKKRRKRESVEQK----- 2463
Db 2135 DKCSPELLSKEDLOJDIORNEIEIEKLEFRVRELEQALLEDRKHFGAVAKBELSLEVOL 2194
Qy 2464 -----RSKONATKLSALFKHKEOLRAEILKKRALLDK-DLOIEVO-----EEL 2506
Db 2195 QAEKRAIDRKEKEITNLEQOL-----EQFRELENNKNEVQOLHMOLEIKKESSTRLOEL 2250
Qy 2507 KRDLKIKKEDLMOLAQATAVAPCPVTPVLPAPRAPPSPRPPOVOTHTLSTPTLP 2566
Db 2251 EENKTLFND-DMEKLGIA-----IKESDMSQODDH 2280
Qy 2567 VASQK-----KRREBKD-----SSSKSKKKMMISTTSKTKDKTKLYCICKPYDESKFY 2617
Db 2281 VLEGFAOIIQOEVEYEDQLENOQVTKLOOOLKITTDNKVIEEKNELIRPLET-----Q 2333
Qy 2618 ICDSDCONMYHGRGCVIIIOSEAEVLIDEVYCPQCSQSTEDAMTVLTPLEKDYGLKRV--L 2675
Db 2334 IEC-----LMSDEQVK-----RNKEEIEOLNEVIEKLOOLANIGOK 2372
Qy 2676 RSLQAHKAMPLEPVPDADYGVVIGIKEPMDLMEERVORRYEKLFEVADMTKIF 2735
Db 2373 TSMNHSLS-----EADS-----LKHQLDVYIAEKLALQOQVETANEETFKKNVL 2419
Qy 2736 DMCXYNPSDSPFYQCAEVLSEFPYOKLGFKASRSHNNKIDS 2778
Db 2420 KETNF-----KMNQLOTELFSLKRESSEVKIOS 2448

RESULT 10
US-09-724-676-68928
; Sequence 68928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68928
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68928

Query Match 2.7%; Score 380.5; DB 5; Length 3907;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 484; Conservative 476; Mismatches 1032; Indels 823; Gaps 116;

Qy 305 DCVAET-----QKNKP--YIRHEPIGYRSRRKYW-----FLNRRLLIE 341
Db 113 DCSSEVNGCSFVMTGKPTNLLRREEFGVDSDYSEGGADSDPTLHLEMSELAGKHCHIE 172
Qy 342 EETDENNEKKIYYTSTKYVLAELIDCL-----DKDYWEAEICLCTL 381

Db 173 ELNRELEEMRNVYTGTEGLOLOEFAAIIKORDCIITOLTANLOQARREKDETMRELELT 232
Qy 382 EEMRE-EIH-RHMDITEDLTNKARGSNKSFIAANEEI-----LESIRAKG 426
Db 233 EOSQKLOIOPOLOASSETLRNSTHSTAAIDLQAKOILTHOQOOLEDHLLIEDYOKKE 292
Qy 427 DI-----DNVKSPEETEKDKNETENDSKDAENRE-----EFEDOSLEKDS 467
Db 293 DFTMQISFIOEKIK-VYEMEDQK-KVENSNNKEIQCKETIIEELNTKIIIEEKKTELDK 350
Qy 468 DDKTPD-----DDEQCKSEYGDFFSKSNGELSESGAG--KASGSTRITRLRNP 518
Db 351 KLTADKLLGELOEQIVQKNOEIKNNKLELTNSKOKEROSSEBIEKIMGTVELOKRNK 410
Qy 519 DSKL-----SOLSQOAAAHAA-NKLFEGKEVLVNVSGEISRLSTKEVI 566
Db 411 DSQFETDIORMEOETORKLEOLRAELDEMYGOIYOMKOBELROMHMAOMEKTRHNKE 470
Qy 567 MKNINNYFKLQOEGKYRVYHNOYSTNSFALNK--HQHREDHDKRHHLAKKCTLPAGER 624
Db 471 MENALRSYNI-----TVNEQIKLMVAINELNIKIDOTNSQKKEKLEEL----- 516
Qy 625 KMGSGVHSGKVLITSLRLITTOLENNITPSFLHPMASHRANWIKAYOMCSKPREFALA 684
Db 517 ---GLTIEEKCALQOLELVEEL-----SESRQ--IQRKQITAEQ--SKLNEAHS 564
Qy 685 LAILECAVAPVWMLPIWREFLGHTRLHRMST-----TEREKEKVKKKEKOEFEETMOOA 740
Db 565 LSTVEDLKAEIYASASRKELELKEHAETVNTYKIKLEMEKEKNAVLDMAESQEBELER 624
Qy 741 TWVYKTPPVKHOWWOKGE--EYRVYTGCGWSWISKTHVYRVPKLPQNTVNVKRSLE 797
Db 625 LRTQLLFHSHEELSKKEDELEIHRIN-----IEKLNKNGIHXYQOID 668
Qy 798 GTKNNDENMD--ESQKRKCSRPKIKIEPDSEKDEKVGSDAKAD-----QNMK 847
Db 669 GLONEMSOKIETWOFEKDMLITKONQILILEISKDLOOLVNSKSEEMTLOINLOKEI 728
Qy 848 DISKITEK-----DQVYKEL-----LDSQSDKCKKEPMEV 879
Db 729 EILROPEKEKGLIEDVOELQKTELLEKOMEKEKENDLOEKROLEAE--NSLIKDEKCTL 787
Qy 880 DDMKTESHVNOQESSQVYVNVSEGFHLRTSYKKKTKSKLDGLERRIKOFTLEEKOR 939
Db 788 EDMLKIHTPVQOE-----RLIFLDSIKSKSDSVWEKEI--ELIEENED 831
Qy 940 LEK--IKLEGGJIKGICKITNSKNL-----SESPVITAKBCCQSDSMROF---Q 985
Db 832 LKOOCIQLENEIEKORNTFSFAEKNEEVNYOELQETACILKYKDLDESKKNOQELEKRS 891
Qy 986 SPNANDOPEDLIDQCSQSDSSVLRMSDPSHTNKKYPRDVLVDYSIRSPT----- 1038
Db 892 KKLALNEE-----LHLQRIINPTTVKM--KSSVFDEKKTVAETLEMBEVY 934
Qy 1039 -----KCPKONSIENTIEEKVSDLASRGQEPKTSKRGDNFTIDSK-LASADD 1086
Db 935 EKQOTELMEKLEVTYKREKLE--LSQRLSDLS---EQLKQKGEISFLNEEYKSLKPEKE 988
Qy 1087 IGTLLCKNKKPLIOESDPTIYSSSKSALHSSVPKSTINDADATPLSRAMPFEKGLCDSDS 1146
Db 989 PVSILRCR-----ELEIINHNR-----ENVOSCDTQV 1016
Qy 1147 NSTLENSDPTVSIODSEEDMIYONSNEISDPOFTRREDQVEVLEPLKCELSGESTGNC 1206
Db 1017 SSLL-----DGVYTMYSRGAGSVSKYKNSFGESKILMVDDKYSF-----NMYGSESKE 1068
Qy 1207 E---DRLP--VGTETANGKPKSQOKKLEERPVNKCSDQIKLNTTDDKNNENRESEKGO 1261
Db 1069 QILDLHLPSVTWESSSLRANQPSENDKLQ-----KELNVLKSQENDLR--LQMEKQ 1116
Qy 1262 R-----TSFQI-----NGKDNKPKIYLGECLEKISESRVYSGNVPRVNNINKTI 1308

```

Db 1117 RICLSIVSTHVQVREYEMENED-----KALCSLKEELIFA--OEKIKELOK-I 1164
QY 1309 PENDISLTYKESAIRPFIINGD-----VIMEDFNERNSETSHLSSSDAGCNRDLSL 1362
Db 1165 HOLELOMKTQET-----GDEGKPLHLILGLOKAVSEESYFIOTLCSUGY----- 1213
QY 1363 ETLPTFRESDSOTOTTPSASCSPESNSVQVEDMEIETSEVKKYTSPTISPIISEESNLSNDF 1422
Db 1214 -----YTPALKC-----EVNAEDKEN-SGDV 1233
QY 1423 IDENGLPIKN-----ENVNGESKRKRYITEVT-----TMTSVANESKTVIVENG 1469
Db 1234 ISNEBPELODORYEVOQDFQENMHITLTKYEEYNKLLVLOTRLSKIMWGQOIGMKLEFG 1293
QY 1470 DKQTVVSTENCASKSVTTTTTIVTKLSTPTSGSDIISVKBOSKTVVTTTTVTDLSITTT 1529
Db 1294 E-----ENLPRK-----ETEFLSHSO----- 1310
QY 1530 GGTLYVTSMTVSKESYTRDKLMLKFSRPRKTRSGTALPSYRKFVTKSKKIFLPPNDL 1589
Db 1311 -----MTYLEDIDVNHKSKLSLODEKTKLEBOVOLESLISSLOQ-----L 1354
QY 1590 KTLARKGIREVPYFNNAKRALDIMPYSPRPTFGITWRYRLQTVKSLAGVSLMRLMW 1649
Db 1355 KETEO-----NYEAE-----IHCLO--KRLQAVS----- 1376
QY 1650 ASLRWDMMAKVPBGSGSTETETSETITTEIIRKRDVPGYGRFGYCRKILICPIGPV 1709
Db 1377 -----ESTVPP-----SLPDSVAVITESDAOQRTMPS-----CYKKNV--DGTI 1414
QY 1710 ETEKPTTPORKGLRSSALRPKRPETPKQTPYLIETWVA--EELFIMETRAAEVREKE 1768
Db 1415 EFSEPEVKKETINIVKLEKOYOQOLEEVAKIVYSKISAFACQOTELSRISGKEMTASS 1474
QY 1769 K-AQAVQQAARKLEQOKPVIVATSTTSPTSSTTSISPAQKVMVABISSVTTGKMWL 1827
Db 1475 KOAHAVCOQBOHYFENEMK-----LSODQIGFOFEIVDVKEFEKFLSKLEHGEKEL 1529
QY 1828 TTKVGSATVTFQONKMFHOTPATWVGOGOSNGVVOQKVLGIIPSSGTSOQPTSF 1887
Db 1530 LS-----NSDPHD-----IPESKCVLTI-----SEMFPSKD 1556
QY 1888 QPRTATVIRPNTSGSGTTSNSOVITGPQIRPMVITRPLQOSTLGKAIIRTPVWOP 1947
Db 1557 KTLFVRSIHDEISVSMDASRLMNEQLEDRQELVROYOEHQOATELROAHMRQM 1616
QY 1948 GARQOVMTQIIRQOPVTAIVAPNTVSTPGOKSLTSATSTSNQSSASQPPRQOQVK 2007
Db 1617 EROHEDEOL--OEETKRLNROLAORSS:IDENLVSEERVLLELEALKOLSLAGREK 1673
QY 2008 LTMALQLOLQGHGNGGLVIVLOGOGTQO-----LOLIPQV--TVLPGRQOQLMQ 2059
Db 1674 LCCELRRSSSTOTONGN-----ENQSEVEEQTFKEKELDRKPEDVPELLSNRYALOK 1726
QY 2060 AAMPNGTVORLETPLATATTAFTAST-----TTTTVSTAATGEOBOSKLSPPMOY 2110
Db 1727 A--NNRLKLILLEVVTATAVEETIGRHVILGILDRSSKSSQSSASLIMRSEAEASVKSVCV 1783
QY 2111 HOKTLP-----PAOSSVGP-----AKAQPTAPPSARPOPQTPQSPAPE----- 2153
Db 1784 HEETHTVDESIPSYSGSDMRNDINMMSKVTEGTELSORLVKSGPAGTEIDPENELM 1843
QY 2154 --VOTQEVQOTVSSHVSEAOPTHAOSSKPOVAOS-----QPSNVQOGSPVAVOS 2206
Db 1844 LNISSRIQAAVEKLEIASETSSOLEHAKVQOTELMRESROKOEATESLKQDELPERL 1903
QY 2207 PSQTRIPSPSOLSPQO-----OSOVQTTT-----SQPIDIQPHTSLQIP 2247
Db 1904 HEESRAREQALVELSKAEGVIDYADEKTLFERQIQEKTIDIDRLQELLCASRLQELE 1963
QY 2248 SOGPOPOVOVOSTOTLSSQOTINQVSSPSRPOLOIQOPQVQVAVAPQLOQOVVLS 2307
Db 1964 AE-----OOQIQEBERELLSR-----QKEMAKAEAP--VBO-----OLQOETKLM 2002

```

```

QY 2308 QIOSOVAAIOAQQSGVPOOIK-LQPLQIOQSSAVQTHQIONV-----VTQAAVQEOQL 2362
Db 2003 KETLEVOCAENKVRDLOQAVALEIDVEQVSRFTELEQENKTELMDLRQONQALEKOL 2062
QY 2363 QRVQOLRDOQOKKQOQOIKREHTLQASNOSEITOK-OVAKHNA-----VIEHLKOKK 2416
Db 2063 EKMRKFLDEQADIREHERHVRFOOELQOLKVRFPQIPISBHQTRVEQOLANHLKEGT 2122
QY 2417 ---SMTPARRENQRMIVQNVKTYILDKID--KEKOA-----AKRKREESV 2460
Db 2123 DKCSLELLSKEDLOQDI--QERNBEIEKLEFRVRELEQALLVSADTFQKVEDRKHFQAV 2179
QY 2461 EOK-----RSKONATKLSALLFKHKEQLRAEILKRALDK-DLOIEV 2502
Db 2180 EAKPELSEVOLOAERDAIDRKKEETINLEQOL--BQFRELEKKNQEVQOLMHQLEI 2235
QY 2503 Q-----EELKRDILIKKEDLMQLOAQTAVAAFCPPVTPVLPAPRAPPSPPPPPGV 2554
Db 2236 OKKESTTRLOELBOENKLEFD-DMEKILGLA----- 2265
QY 2555 QHTGLSTPTLPVYASOK-----RKREERD-----SSSKKKKKMSTSTSKETKDTKLYC 2605
Db 2266 KESDAMSTODQHVLFGRFAOIIQEKVEEIDQLENOVYTKLOQOLKTTDNKKVIEKNELIR 2325
QY 2606 ICTPYDESKFYIGCDRCQNMVHGRVGLIOSEAEILIDEVYQPOCOSTEDAMTVLPTE 2665
Db 2326 DLET-----QIEC-----LMSQOECVK-----RNREBELQLENEVIE 2357
QY 2666 KDVEGLKRV--LRSLOAHKAMPLEPVPDNDAPDYGVYIKEMPDLATHEERYOKRYRK 2723
Db 2358 KLOQELANIGOKTSMANHSLS-----EADS-----LKHOJDVIAIKLALQEOVET 2404
QY 2724 LTFEYVDMTFKIPNCRVYNSDSPFQCAEVLESFVOKLKGFKASRSNNKLOS 2778
Db 2405 ANEEMTFMKNVLETFN-----KKNQLOTELSLKRRESVEYKIOS 2445

RESULT 11
US-09-724-676-68916
; Sequence 68916, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 68916
; LENGTH: 3919
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-724-676-68916

Query Match 2.7%; Score 380.5; DB 5; Length 3919;
Best local Similarity 17.2%; Pred. No. 0.017;
Matches 484; Conservative 476; Mismatches 1032; Indels 823; Gaps 116;

QY 305 DCVAET-----QKNKP--YIRHEPIGYRSRRKTV-----FLNRLLIE 341
Db 125 DCSSEVNGCSFVNRGTGPTNLREEREGVNDYSQGAQDSPNLMMESELAKGQHELE 184
QY 342 EDRENEKEKIWTYSTVOLAEILDCI-----DKDYWEAEILCKIL 381
Db 185 ELNRELEKRVYTGTEGLOLOQEFBAIKORDGIIITQLRANLQOARREDETFMRHELELT 244
QY 382 EEMRE-EIH-RHMDITEDLTKNARGSKSEFLAANEI-----LESIRAKKG 426
Db 245 EOGOKLQIQFOLOASFTLNSHSSVYAADLLOAKQOILTHQOULEBODHLLLEDYOKKE 304
QY 427 DI-----DNVKSPEETBKDNENENDSKQAEKNRE-----EFEQSLKKS 467

```


Db 305 DFTMOISFIOEKIK-VYEMQODK-KVENSNKBEIOEKETIIIEBLNKKIIEEKKITIELKD 362
Qy 468 DKTPD-----DDEQOKSEVGDPKSEKNGELSESPGAC--KGASGSTRITIRLRNP 518
Db 363 KLTITADKLIGELOEQIVQKNQOEIKNNKLELTJNSKQERQSSSEETKQIMQVLELOKRNKH 422
Qy 519 DSL-----SOLKSQVAAAHBA-NKLEKGEKVLVNSQGEISRLSTKEEYI 566
Db 423 DSOFEFDIVQRMEOETQORLEQIRALDEMYGQOIVQMKOELIRQMAOMEKKITHKGE 482
Qy 567 MKNINNYFKLOGEGKYRYVHNOYSTNSFALNK--HQHREDHXKRHHLHAKFCOLTAGER 624
Db 483 MEALMSYSNI-----TYNEDOIKLMAVAINELNLIKIDDTNSQKEKLEEL----- 528
Qy 625 KNGSVYHSGKVLITSLRLITOLENNIPSSFLHPNMAASHRAMIKAVOMCSKPREFALA 684
Db 529 ---GLILEEKCALQROLEDLVEEL-----SFSRQ--IQRAQOTIABDE-SKYNLAHNS 576
Qy 685 LALIECAVAPVWMLPIWREBELGHTLRHMTS-----IEREKEKVKKKKEKQEEETMOQA 740
Db 577 LSTVEDLKAIEIVASBSERKELELKEAEVTNYKIKLEMEKEKNAAVLDRABESQEALEER 636
Qy 741 TWVKYTFPKHVOWKQGE---EYRVYGYGMSWIKTHVYRFPRLPGNTVNYKRSLE 797
Db 637 LRTQOLLFSHEEELSKLEDELETHRIN-----IEKLDNGLGIHYKOQID 680
Qy 798 GTRNNMDND--ESDKRRCSPRPKIKIEPDEKDEYKSDAANKGAD-----QNEH 847
Db 681 GLONEMSKOETIMQFEDKDLITKQNLJIEISKLDLOQSLVNSKSEBMTLOINELQKEL 740
Qy 848 DLSKITEK---DDVYKEL-----LDSDSKPKCEEPMEY 879
Db 741 ELTROBEKEKGTLEQVOELQKTELLEKQMKENKDLQEKFAOLBAE-NSILKDEKCTL 799
Qy 880 DDDMKESHVNCQESSQOVVWVNSEGFHLRTSYKKKTKSKLDGLLEHRRIKQFTLEKOR 939
Db 800 EDMKLTHTVSOE-----RLIFLDSIKSKSDSVWEKEL-ELLIBEND 843
Qy 940 LER--IKLEGGIKIGIKTJNSKNL-----SESPIYTAKEGQSDSMROE---Q 985
Db 844 LKQOCIQLEBIEKQNTSPFAEKNFEVWQELQEEYACLLAKDLEDSKKNQOLEYKS 993
Qy 986 SPANNDOPEDLIGCCSOSDSSVLRMSDPSHTTNKLYPKDRVLDDVSIKSPET----- 1038
Db 904 KKLALNEE-----LHLQIRINPTTVKM--KSSVDEDEKTVYAEITLMEGEVY 946
Qy 1039 -----KOPKONSTIENDEEKVSDLASGOEPTSKIRKGNDFIDDSK-LASADD 1086
Db 947 EKOTTELMKELEVTKREKLE--LSORLSDLS---EOLKQKPEISFLMEEYVSLKPERE 1000
Qy 1087 IGTLLCKNNKPLIOESDPIVSSSKSALHSSVPKSTNDRDATPLSRAMPFECKLGDSES 1146
Db 1001 PVSILRR-----ELEIINHNR-----ENVQSCDTQV 1028
Qy 1147 NSTLENSDPTVASIODSEEDMIVQNSNESISQOFTREDOVEVLEPLKCELVSGESTGNC 1206
Db 1029 SSL-----DGVVTMTSRGAEQSVSKVNSFGESKIMVDXVSFE---NMFTGGESEQE 1080
Qy 1207 E---ORLP--VAGTEANGKKPSQOKKLEERPYNKQSDQIKLANTTDKKNENRESKQO 1261
Db 1081 QILDLHLPSVTRESSLRATQPENDKLO-----KELNVLKSEONDR-LQMEAO 1128
Qy 1262 R-----TSTFOI-----NGKDNKPKIYLKGECLKEISERVYSGNVEPVNINNKII 1308
Db 1129 RICTSLVSTHVDQVREYENKED-----KALCSLKEELIR--OEERKIKELQK-I 1176
Qy 1309 PENDIKSLTVKESAIRPITNGD-----VIMEDFNERNSSETKSHLSSSDAEGNROSL 1362
Db 1177 HQLEIQTKMTQET-----GDEGKPLHLLIGLKQKAVSEESYFLQITCSVLGER----- 1225
Qy 1363 ELLPSTKESDSTQTTTPASCSPEVSNVNOVEMEIETSEVKKVTVSPTISEESNLSNDF 1422
Db 1226 -----YTPALKC-----EYVAEDKEN-SCDY 1245

Qy 1423 IDENGLPIRINK---ENVNESEKRRVTITEVT-----TMTSTVATESKTVIKERKG 1469
Db 1246 ISENEDEPQLODYRAYQODPOENHMTLLNKVTEEYKNLVLQITRLSKIMWQOQTGKMLERG 1305
Qy 1470 DKQTVVSTENCAKSTVTTTTTIVTKLSTPBGSGVDIISVKQKSTVVTVTYDLSLTTT 1529
Db 1306 E-----ENLPEK-----EPEFLSIHQ----- 1322
Qy 1530 GGLTVMSTVSKESTRDVKLMLKFRPKKTRGCTALPBYRKRVTSTKRSIFVLBNDL 1589
Db 1323 -----MTNLEDIVNHSKSLSSLODLEKTKLEBOVOLESLISSLOQ-----L 1366
Qy 1590 KKLARKGIREVYFYFNNAKPALDIMPYSPREFTGITWRRLQTVKSLAGVSLMLRLM 1649
Db 1367 KETEQ-----NYEAE-----IHQO--KRLQANS----- 1388
Qy 1650 ASLRMDMAKYPGGGSGSTRETSETEITTTETIKRRDVGPIGIREYCIKRTICPIGV 1709
Db 1389 -----ESTVPP-----SLPVDVSVITTESDAQRTWPGS-----CVKKNI--DGTI 1426
Qy 1710 EPEKETPTQORGLRBSALRPRKPEPTPKQGPVILETWY-BEELELMEIRAEAEVKE 1768
Db 1427 ERSGEFVKEETNIVKLEKQIOBQLEEBEYAKYVMSIAFAQOTELSRISGKENTASS 1486
Qy 1769 K-AQAVEQOAKKRLQEOOKPVIATSTTSPTSTSTISPAOKVWVAPISGSVYTGKMYL 1827
Db 1487 KQAHAVQOQOEHFNEMK-----LSQDQIGFOFTEFVDVKFKFEFRLKELGEHKEKL 1541
Qy 1828 TTKVSPATVTFQOONKNPHQTFATWVKQOOSNGVVOQKVLGITPSTGTSQOFTSF 1887
Db 1542 LS-----NSDPHD-----IPESKDCVLT-----SEEMFSKD 1568
Qy 1888 OPRATVTRIPNTSGSGGTTNSQVITGPQIRGMYVITPILQOOSTLKGALIRTPMYNP 1947
Db 1569 KTFIVQOSIHDEIVSSMDASROLMLNEBQLEDMQELVRQIOEHQATELLQAHMRQ 1628
Qy 1948 GAPQOVMTOIIRQPVSTAVSAPNTVSPGOKSLTASVSTNSIQSSASQPPRQOQV 2007
Db 1629 ERQREBOBL---QEIRIKLNRQLAQRSSIDNENLVSENERVLEBELLEMLKOLSLAGREK 1685
Qy 2008 LTMALQTLQGHGNGOGLTVVIOGGQTTGQ-----LQILPGV--TVLDEPGQOLMQ 2059
Db 1686 LCCELNNSSTQONGN-----ENQGEVEEQTFKEKLEDRKEDVPPELISNERVALQ 1738
Qy 2060 AAMPNGTVORLEFTPLATTATTAFT-----TTTTYSTTAAGCEOROSLSTQOMY 2110
Db 1739 A--NNRLKILLEVVKTAAVEETIGRHVGLIDLRSSKSQSSASLIMRSEAEASVSCV 1795
Qy 2111 HODKTLR-----PAOSSVGP-----AKAOPQTAQPARPQTPQPSQAPQ----- 2153
Db 1796 HEHTVNTDESIPSTYSGSDMFRNDIMMSKVIEEGELSQRLVRSFGAETIDEPEHELM 1855
Qy 2154 --VQTOPEVQOTQTVVSHVPSAEAPTTHAOSKPOVAOS-----QPOSNVQOSPVYOS 2206
Db 1856 LNISSRLQAAVERLEBAISETSSOLEHAKVQTELMKRESFRQKEATBESLKQOELEREL 1915
Qy 2207 PSQTRIRPSTPSQSPQ-----OSOVOTT-----SQPIPIQPHSLQIP 2247
Db 1916 HESRARQOLAELSKAEVINDGAYADEKTLFERQIOEKTDIIDRLQOEELLCASNRLOELE 1975
Qy 2248 SOGQPOPOVOVOSTQTLSSGOTLINOVSVPSPRPQLOIQOPQVIAVAPOLQOQVUS 2307
Db 1976 AE-----QOQIOBERELLR---QKNAKAEKGP--VEQ-----QLQOETKLM 2014
Qy 2308 QIOSQVVAOIQAOOSGVPOIK-LQLPQIOOQSAVOTHOIONV---VTVQAAVQEOUL 2362
Db 2015 KEKLEVOQCAKEVPRDLQVQKALEIDVEQVSRFTELEQKNTLMMDLRQONQALEKL 2074
Qy 2363 QRYOQLRDQOQKKQOQOIEIKREHTLOASNSEIIOK-OVYKHNH-----VIEHLKOKK 2416
Db 2075 EKMRKFLEDOAIDREHERVDFOEIOKLEOOLKVVRBFRPISIEHOTREVEOLANHLKERT 2134

QY 305 DCVAEI-----QKNP--YIRHEPIGDRSRKYW-----FLNRRLIE 341
 Db 113 DCSEVNGCFVNRKGTPTLRLREERGVDVSYSEGAQDSPHLEMESELAGKQHEIE 172
 QY 342 EDENENEKIMWYSTRVQLAELIDCL-----DKDYWEALCKIL 381
 Db 173 ELNRLEEMEVYVTEGLOLOEFAAIKORDGIITOLTANTLOQARREKDETMEEFELT 232
 QY 382 EEMRE-BIH-RHMDITFDLNRKAGSKSFLAANEI-----LSISRAKK 426
 Db 233 ESOCKLOLOFOQLOASTLKNSTHSTAADLOAQOOLTHQOOLEBODHLEBYOKKE 292
 QY 427 DI-----DNVKSPEETEKDNETENDSKDAEKRE-----EFEDOSLEKDS 467
 Db 293 DFMQOISFLOEKIK-VYEMQDK-KYENSKNEKEIOEKETIIEELNTKIIEEKKITLEKD 350
 QY 468 DDKTPD-----DDPEQSGSEVDKSEKSNGBLSESPAG--KGASGSTRITTLRNP 518
 Db 351 KLTTADKLGLGLOEOLVOKNOEIKNMKLELTNSKOKEROSSEIKQIMGTVEELOKRNHK 410
 QY 519 DSKL-----SOLKSOYAAAHAA-NKLFKEGKEVIVYNSGCEISRLSTKREVI 566
 Db 411 DSOPEDIYORMEQETORKLEOLRAELDEMYGOQIYOMKOLIKOHNAQHEMKTTRKGE 470
 QY 567 MKNINNYFLOGEGKYRYVHNOYSTNSFALNK--HOHRDHDKRRLAHKFCITPAGER 624
 Db 471 MENLRYSNI-----TVNEDOIKLMMNVAIINELNIKTLOJNSQEKLEEL----- 516
 QY 625 KMGNSVSGKVLITSLRLITTOLENNIPSSFLHPNASHRAKIVQMSKPREALA 684
 Db 517 ---GLILEECALOROLEDIVEEL-----SFSREO--IGORATIAEQ-SKLENAHS 564
 QY 685 LAILECAVKPVVMLPIWREFLGHTRLHRMTS---IEREKEKVKKKKEQOFEETQOA 740
 Db 565 LSTYEDAKAELIVSASESRKLELKEHAEVYNYKIKLEMLERKNAVLDRAESEAELER 624
 QY 741 TWKYTFPVKHQWKQGE---EYRVTYGSGWSISKTHVRYEVPKLPNGTNNVYRSLE 797
 Db 625 LRTOLFSHEEELSKLEDEIEHRIN-----IEKLDNIGIHYKOID 668
 QY 798 GTKNMDEMD--ESDKRKSRSRPKIKIEPDSKDEKVSDAKAGD-----ONEM 847
 Db 669 GLOKESOKLETWOFEKDNLTIKOMOLILEISKLDLOQSLVNSKSEMTLOJINLOKEI 728
 QY 848 DISKITEKK--DQDYKEL-----LDSDSKPCKEEPMY 879
 Db 729 EILROKEKEGTLEQEVQELQKTELLEKQMKKEKENDLOEKFAOLEAF-NSILKDEKTL 787
 QY 880 DDDKTESHAVNCOSSQOVAVNVSEGFHLRTSYKKKTSSKLDGLERIKQFTLEEKOR 939
 Db 788 EDMKIHTPVSQEE-----RLIFLDSIKSKSDSVMEKEI-EILIEENED 831
 QY 940 LEK--ILLEGIGIGICTSNSSKNL-----SESPVITTKAKEGCOSQSMRQE--Q 985
 Db 832 LKOOCLOLNEIEKORNTFSEFAEKNEFVNYQEOLEEYACILKYADLLEDSKNOQOLEYKS 891
 QY 986 SPANNDQPEDLLIOGCSQSSSVLRMSDPSTHTNKILPKRVLDDVYSIRSEP----- 1038
 Db 892 KLAALNE-----LHQRINPTIYKM--KSSVPEDEKTPFAETILEMEVY 934
 QY 1039 -----KCPKONSINDIEEKVSDIASRGOEPTKSKTKGNDFIDSK-LASAD 1086
 Db 935 EKDTTELMERLEVYKREKLE--LSQRLSDLS---EQLOKQKPGISFLNEEYVSKLPEKE 988
 QY 1087 IGTILCKNKKPLIOESDITVSSSKSALHSSVPRKSTNDROTATPLSRAMDPEGLIGCSES 1146
 Db 989 PVSILRC-----ELEIILHNRA-----ENVQSCDQY 1016
 QY 1147 NSTLENSDVTYIODSEEDMIYONSNEISIEOPRTREDOVEYLEPLKCEVSESGCNC 1206
 Db 1017 SSLL-----DGVYMTSGAGAGSVSKVKSFGESKIMVEKVSFE---NMTYGEESKOE 1068
 QY 1207 E---DRLP--VKGTENGKRPQOKKLEERPVNKCSQOIKLKTNTDKKNNESESEKGO 1261

Db 1069 QULIDHLPSYTKRESSLRATQPSENDKLO-----KELANVLSKQNDLR-LQWEAQ 1116
 QY 1262 R-----TSTQI---NGKDNKPKIYLKGBCLKEISEKSVSGNVPKYNNINKIT 1308
 Db 1117 RICLSLYSTHVDVOREYEMENEKD-----KALCSLKEELIFA--QSEKIKLEOK-I 1164
 QY 1309 PENDIKSLVYKESAIRPFINGD-----VIMEDFENERSSETKSHLLSSDAGNYDNL 1362
 Db 1165 HOELQIMTKQET-----GDEBKPLHLILGLKQKAVSECSYFLOTLSVIGEY---- 1213
 QY 1363 ETLPSSTKESDSTOTTPSPASCPSNSVNOVEDMETSEVKVYTSPTITSEESNLSNDF 1422
 Db 1214 -----YFPALC-----EVNAEDKEN--SGDY 1233
 QY 1423 IDENGLPIKNN---ENVNGSKRKYITVEY-----YMTSYATSKIVYIYERK 1469
 Db 1234 ISENEDELQDYREYVODFENHMTLLNKYTEEYKNLLVLQTRLKISWGOQDGMKLEFG 1293
 QY 1470 DKQTVASTENCAKSTVTTTTVTYTKLSTPSTGSDVLIISVKEQSKTVVTTVTDLSLTT 1529
 Db 1294 E-----ENLPKE-----EFELSIHQ----- 1310
 QY 1530 GGLVTSMTVSKSEYSTRDYKYLKFSRPKKTRSGTALPSYRKVYVSTKKSIFVLPNDL 1589
 Db 1311 -----MTNLEDIDVNHKKSLSLOLEKTKLEQVOELESLSLOO-----L 1354
 QY 1590 KILARKGIGREVPFNFNARPAIDWYPPRPFTGTMTMYRLQOTYVSLAGSILMLRLM 1649
 Db 1355 KETBO-----NNEAE-----THCLO--KRLQAVS----- 1376
 QY 1650 ASLRMDMAKVPBGSGSTRETSETIETTLIKRDNVPGYIGRFEYCIARKIICPIGPV 1709
 Db 1377 -----ESTVPR-----SLPVDVVLTESDAQTMVPGS-----CYKNNI--DGTI 1414
 QY 1710 ETRKETPTPOQRKGLRSALPKRPETPKQTPVLIETWA--EELIEMELRAAERYEKE 1768
 Db 1415 EFGSEFVKKEETNIVKLLLEQYOOLEEYAKIVYSMSIAFOOTELSRISGKEKENTASS 1474
 QY 1769 K-AQAVQOAKRRLQOQKPIVYATSTSPSTSTSPAQVAVPISGVTGTGKML 1827
 Db 1475 KOHAHVQOQOHYFNEKK-----LSQDQIOGFTEYDVYFKKEFKPLSELGHEKEL 1529
 QY 1828 TTKVSPATVTFQONKNFOTFATVWYKQSGNSGVQOVLGIIIPSSGTSGOQFTSF 1887
 Db 1530 LS-----NSDPHD-----IPSKOVLTI-----SEBMSKD 1556
 QY 1888 QPRTAVYTIKPNISGSGGTTNSGOVITGPQIRPMYVIRPILQOSTIGKAIIRTPVWOP 1947
 Db 1557 KTFIVROSIDEISVSSMDASROIMLNEQDE-----DMROE-----LVRO 1597
 QY 1948 GAOQVWTOITIRGQPVSTAVSAPRTVSTPGOKSLTSATSTNSIQSSASQPPRPOQGVK 2007
 Db 1598 YOEHQATELDR-----QAHMROMEKROEQ 1625
 QY 2008 L--TMAOLT-OLTHGSG--NOGLTVVIOGOGOTTGOLIPQV--TVLPBGQOOLMOA 2061
 Db 1626 LOEETIKRLNLQAOQNENQO---EVEEQTFKKEKLEDRKPEDVPRILLSENRALOKA- 1681
 QY 2062 MPNGTVOREFLFTPLATTATAST-----TTTVYSTTAAGTEGROSKLSPOQVHQ 2112
 Db 1682 --NNRLKILILEVYKTTAAEETIGRHVLGILDRSSKSQSSASLIMNSEAEVASKCVHE 1739
 QY 2113 DKTLIP-----PAOSSVGP-----AKAOPQIAQPSARQPOQPOPSAPOPE----- 2153
 Db 1740 EHTFVDESIPYSGSDMPNDINMMSKVTEEGTELSQRLVRSFGAGTEIDPNEBELMLN 1799
 QY 2154 VOTOPEVOTQTVVSHVPSAOPTHAOSKPOVAAS-----OPOSWQOSQSVRYQSPS 2208
 Db 1800 ISSRLQAAVAKLELAISSETSOLEHAKVYOTELMRESFRKQKQATELSKQOELREKHE 1859
 QY 2209 QTRIPSTPSQSPQO-----OSOVQOTT-----SQPIQOPHTSLOIPSO 2249


```

Db 1165 HOELQTMKTQET-----GDEGKPLHLIGLQKRAVSECSFYLOTLCSLGEY---- 1213
QY 1363 ETLPSKESDSTQTTTSPASCPESSNVNQVEDMEIETSEVKVYTSSPITSEESNLNDF 1422
Db 1214 -----YTPALKG-----EVNAEDKEN-SGDY 1233
QY 1423 IDENGLPINKN---ENVNGESKRRTVITEVT-----TMTSTVANEKSTVIVKRG 1469
Db 1234 ISENEDEPLQDYRYEVDQFENMHITLTKVTEYENKLLVQTLRLSKIMWGQOTGMKLEFG 1293
QY 1470 DKQTVVSTENCASKSTVTTTTVTVKLSTPSTGSDVLIISVKQSKSTVTTVTDSLT 1329
Db 1294 E-----ENLPKE-----ETEFJLSHSQ----- 1310
QY 1530 GGLVTSMVTSEYSTRDKVLMKFSRPKKTSGTALPSYRKPVTKSTKSIPLPNDL 1589
Db 1311 -----MNLDEDIVNHSKSLSDLEKTKLEBOVQELSLSSLOO-----L 1354
QY 1590 KLIARRGIGREVPYFNNAKRALDIMPSPRPFTGITWRYRLQTVKSLAGVSLMRLW 1649
Db 1355 KETEQ-----NYEAE-----IHCLQ--KRLQAVS----- 1376
QY 1650 ASLRMDMAKVPYPPGGSTTETSETETTELKRRDVPYGIREFYCIKRIICPIGP 1709
Db 1377 -----ESTVP-----SLPVSVTITESDAQRTMYPGS-----CYKKNI--DGTI 1414
QY 1710 ETPKETPTPOKRLRSSALRPKRPETPKOTGPYIETWVA--EEELMELTIRAEAREYKE 1768
Db 1415 EFGEGEVKKEETIVKLLQYOBQLEBEAKIVSMTSIFAQOTELSTRISGKENTASS 1474
QY 1769 K-AQAEQQAQRRLQEQKPTVATSTTSPSTSTISPAQKVAVAFISGVTGTGKMWL 1827
Db 1475 KOAHAVCOQEOHAFENEMK-----LSODQIGQFTEFYDVYKFEKPKPLSKELGHEGKEIL 1529
QY 1828 TTKVGSBPATVTPQONNENHOTFATWVKQGSNSGVQVQKVLGIIPSSGTGQQTFS 1887
Db 1530 LS-----NSDPHD-----IPESKCCVLT-----SEEMFSKD 1556
QY 1888 QPRTATVTIRPNTSGSGGTTNSQVITGPIRPGMTVIRTPLOOSTLGAIIPTPVVCP 1947
Db 1557 KTFIVQSHIDEISVSSMDASRLMNEBLE-----DMQOE-----LVRO 1597
QY 1948 GADQVWQTLIRGQVPYSTAVSADPNVSSPFGOKSLTSANSTSNISQSSAGPPPPQGGQVK 2007
Db 1598 YQEHQOATELLR-----QAMHRQMERQREDOEQ 1625
QY 2008 L--TMAQLT-QLTQGHG--NOGLTVIYQOGQTTQLOLIPQV--TVLBPQOQLMQA 2061
Db 1626 LOEIRKRLNQLAQONGENQOQ--EVEBOTFEKELELDRKRPEDVPELISNERVALOKA- 1681
QY 2062 MPNGTVQRFLEPLATTATTAFT-----TTTVSTTAAGTGBOROKLSLPOMQVHQ 2112
Db 1682 --NNRLKILLEVYKTTAAVEETIGRHVLGILDRSSKSSQSSASALIMRSEAEASVKSCHYE 1739
QY 2113 DKTLF-----PAQSSVGP-----AKAQPOTAPSAAPQOTOPQSPQAPPE----- 2153
Db 1740 EHTRYVDESIPYSQDMPRNDIMNMSKVTEEGTELSQLRVSGFAGTEIDPENNELMN 1799
QY 2154 VQVQPEVQVQTVYSHVPSAOPTHAQSSKPOVAAS-----QPSNVQOQSVYRQVQSPS 2208
Db 1800 ISSRLQAAVEKLEAISSETSSQLEHAKVYOTELMRSEFRQKQOATELSLQOEELREHLE 1859
QY 2209 QTRIRSTSTQSLSPGO-----QSOVQVTT-----SOPPLIQPHLSLOIPSQ 2249
Db 1860 ESARARQLAVELSKACGVIDGVADEKTLFERQIQEKTIDITRLQELCLCSNKLQELAE 1919
QY 2250 GQPSQPOVQVSSQTQLSSQOTLQVSVSSPSRQLOIQOPQVIAVLPOLQOQVQVLSQI 2309
Db 1920 -----QOQIQOEBRELLSR--OKEAMKAEAGP--VEQ-----QLQETEKLMKE 1958
QY 2310 QSOVVAQIQAGSGVPOQIK-LQPLQIQOASSAVQTHQIQV-----VTVQAAVQOQLOL 2364
Db 1959 KLEVOQAEKVRDQLQOKVAKALEIDVEOVSRFIELEQEKTELMDLRQONQALEKOLEK 2018

```

```

QY 2365 VQOLRQOQKQKQOQOQLEIKREHTLOASMOSEIIOK-QVAKHNA-----VTEHLKQKK-- 2416
Db 2019 MKREFLQDQALDRBHERVRFQGEIQKLEQOLKVPYPRQISEHQOTREVEQLANHLKEXTDK 2078
QY 2417 -SMTPAERENQMT-VQNOVKYILDKIDKEKQOAKKKRRESEYQK----- 2463
Db 2079 CSELLLSKEQLOIDERNEIEKLEFRVRELQALLQEDKHGVAEAKSLEVOLOA 2138
QY 2464 -----RSQONATKLSALFLPKHEQOLRAEILKKRALIDK--DLQIEVO-----BELKR 2508
Db 2139 ERDAIDKREKETINLEBOL-----EQFRELEKNNEVOQLHMOLEIQKSTTRLODELO 2194
QY 2509 DLKIKREKIDLOLAQATAVAAPCPVTPVLPAPBAPPPSPPPGVOHTGLSTPTPLPA 2568
Db 2195 ENKLFND-DMEKIGLA-----IKESAMSTQPOQHVL 2224
QY 2569 SQK-----KRREBKD-----SSSKSKKKMMISTTSKETTKDQTLVCLCTPYDESKFYIG 2619
Db 2225 FGKFAQIIQKEVEIDOLNBOVTKLOQOLKITTDNKKVIEKKNELIRDLFT-----QIE 2277
QY 2620 CDRQOMMYHGRGVGILQSEAEILDEYVCPQCSQSTEDAMTVLPTLEKDYGLKRV--LRS 2677
Db 2278 C-----LMSDOECVK-----KRREBETOLNEVTEKLOQELANIGQKTS 2316
QY 2678 LOAHKMAPLEVPDNDADYGVVIEKPMDLATMEERYORRYEKLTEVADMTKIFDN 2737
Db 2317 MNHSLIS-----EADS-----LKHQLDVYIAELALQEQVETANEMFTMKVNLKE 2363
QY 2738 CRYNPSDSPFYQCAEYVESFPYQKLGKFGKASSHNNKLOS 2778
Db 2364 TNF-----KMNQLOTELFLSKRRESEVETIOS 2390

RESULT 15
US-09-724-676-68913
; Sequence 68913, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68913
; LENGTH: 3852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68913

Query Match          2.6%; Score 379; DB 5; Length 3852;
Best Local Similarity 17.2%; Pred. No. 0.018;
Matches 481; Conservative 477; Mismatches 993; Indels 850; Gaps 118;

```

```

Db 351 KLTJADKLLGLOEQIVQNOEIKNNKLELTNSKÖKEÖSSEIKOLMGTVELOÖRNEK 410
Oy 519 DSKL-----SOLKSQOAAAHAAH-A-NKLFEGREVLVNSOGELISRLSTKREVI 566
Db 411 DSOFETDIYORMEQÖFQRLQELRAELDEWGOQIVQMOEILROMAQMEEKTRHKE 470
Oy 567 MGNINNYKLOEGKYRYHNOYSTNSFALNK--HÖHEDHDKRHLAKFCLTPAGEF 624
Db 471 MEWAALSYNI-----TYNEDOIKLMAVYAINELNIKLODTYNSÖKEKLEEL-----516
Oy 625 KNNGSVHSGKVLITSLRLITOLENNIPSSFLHPWASHRAWIKAYOMCSKPREPALA 684
Db 517 ----GILIEKCALÖRLELVEEL-----SPSREQ--IÖRÄÖTTAÖEB-SKLNFAHKS 564
Oy 685 LALIECAVAPVYMLPIWREFLGHTLRHMTS----IEREKEKVKKKERKÖEEETMOA 740
Db 565 LSTVEDLKAIEIVASASEKRELEKHEAVTNYKIKLEMEKEKNAYLDMAESQEAELER 624
Oy 741 TWVKTFPYPKHQWOKGE---EYRTYGGGMSWISKTIVYRFVPRPGNTVNYKRSLE 797
Db 625 LRTÖLFFSHEELSKLEDELEIHRIN-----IEKLDNGLGTHYKQOJD 668
Oy 798 GTKNNDENMD---ESDKRCSRSPPKIKIEPSEKDEVKSDAKGAD-----QNM 847
Db 669 GLÖNEMSÖKIEFMÖPEKDNLITKÖNÖLIEISKLADLOÖSVNSKSEEMTLOINELÖKEI 728
Oy 848 DISKITEKK---DÖDVKEL-----LDSDDKPCKEPEMEV 879
Db 729 EILRÖBEKKGLEQVÖLQELTELLEKÖMEKENDLÖEKFQÖLEAE-NSILKDEKTL 787
Oy 880 DDDMKTESHVNOESSÖUVVNVSEGFHLRTSKYKKTKSKKIDGLIERIKÖFTLEKÖR 939
Db 788 EMLLITHPVSOE-----RLIFLDSIKSKSDSWEKEI-ELLIEENED 831
Oy 940 LFK---IKLEGGIKGIGKTYSTNSKNL-----SESPVITAKECQÖDSMOE----Q 985
Db 832 LKQÖCLOANELEKÖRNTSPFAKNEFVYQÖLEIYACLLVYKDLDESKKNQÖLEYKS 891
Oy 986 SPANANDOPEDLLOCSÖSÖSVLRKMSDPSTNTKLPYKDRYLDVYSIRSPET-----1038
Db 892 KKLALNEE-----LHLÖRINPTTVKM--KSSVFDEKXTVAELFEMGEVY 934
Oy 1039 -----KCPKÖNSTENDIEEKVSDLASRQÖPTSKTKGNDFTIDSK-LASADD 1086
Db 935 EKDÖTLEMEKLEVTREKLE--LSÖRLSDLS---EÖLKÖKPELISFLNEEYKSLPEKE 988
Oy 1087 IGTLLCKNNKPLIOESDPTVSSSKSALHSVYKSTNDRDAPPLSRAMFEGKLGDSKS 1146
Db 989 PYSKLGR-----ELEIILHNRA-----ENVOÖCDTÖV 1016
Oy 1147 NSTLENSDVTYSIÖDSSEEDMIVONSNEISÖQFTRQÖDVEVLELKCELVSGESTGNC 1206
Db 1017 SSSL-----DGVTMTSRGEGSVSKVSKFGEESKIMVDKVSFE--NMIVGEESKÖE 1068
Oy 1207 E---DRLP--VAGTEANGKPKPSÖQKKLEERPYNKCSÖQIKLKNYTDKKNENRESEKQ 1261
Db 1069 ÖLILDLPSVTRESSLRATQÖPÖENDKLQ-----KELNVLSKSEÖNDR-LÖMEAO 1116
Oy 1262 R-----TSTFOI-----NGCKNKPKYTLKGECLKEISESRVYSGANPEPVNNINKII 1308
Db 1117 RCLSLVSTHVDÖVREYENKED-----KALCSKEELIFR--OEERIKELÖK-I 1164
Oy 1309 PENDIKSLVKEASIRPFINGD-----VIMEDFNERNSSETSKSHLSSSDAEGNVDSL 1362
Db 1165 HQLELOQTWKTÖT-----GÖEGKPLHLLIÖKLOKAVSECSYFLOTLCSVLGEF----1213
Oy 1363 ETLPSKESDSÖTÖTTPASACPESNSVNOVEDMEIETSEVKVYTSPTISEESNLNDP 1422
Db 1214 -----YTPALKC-----EVNAEDKEN-SGDY 1233
Oy 1423 IDENGLPINKN---ENVNGESKRTVITEVT-----TMSTVATESVTKIVKEVG 1469
Db 1234 ISENDEPÖLÖDRIVÖDQÖENMHLLNKVYEEYKNLLVLOTRLSKIMQÖQTÖGMLKEFG 1293

```

```

Oy 1470 DKÖTVVSTENCARKSTVTTTTVTKLSPTSGSVNDISVKÖSKTVVTVVTSLSITTY 1529
Db 1294 E-----ENLPE-----LEPFLSHSQ-----1310
Oy 1530 GGLTVMSVSKYSTRDYKYLKMFSPKTRSGTALPSTRAKTVYTSKYSIVLPENDL 1589
Db 1311 -----MNLIEDIVNHKSLSLODEKTKLEBOVOELESLSILOO-----L 1354
Oy 1590 KKLARKGIREVYENYNNKPALDIMPSPRPTPTITMRYRLÖTYKSLAGVSLMLRLM 1649
Db 1355 KTEQ-----NYEA-----IHCLÖ--KRLÖAVS-----1376
Oy 1650 ASLRMDMAKYPGGGSGSTRSETSEIITTELIKRRDGYPIGIREYCIKILICIGVP 1709
Db 1377 -----BSTVPP-----SLPVDSVITTESDARMTYPS-----CYKKNI--DGTI 1414
Oy 1710 ETPKEPTPQKGLRSSALRPKRPTPKÖTPYIETWA--EELELMETIRAFAEVEKE 1768
Db 1415 EFSGEFVKEETNIVKLEKÖYÖLEEVAKYVMSIAFAQÖTBLSRISGGKENTASS 1474
Oy 1769 K--AOAVEOAKKRLBOÖKPTVATSTPSSTSTSLSPAÖKVMVAPIGSVTTGKNVL 1827
Db 1475 KQAHAVCÖÖBÖHYFENEMK---LSÖDQIGFÖTFEVTVKFEKFRPLSKEGEHKEITL 1529
Oy 1828 TTKVSPATVTFQÖNKNFQÖTFATWVKÖGSNSGVVQVÖÖKVLGIIPSTGTSQÖFTSF 1887
Db 1530 LS-----NSDPHD-----IESKDCVILTI-----SEMFNSD 1556
Oy 1888 ÖPRTATVTRPMTSGSGGTTNSÖVITGPÖIRPGMTVTRPLOÖSTLGAIIITPYVOP 1947
Db 1557 KTFIVRÖSHIDEISVSMASROLMEQLE-----DMKÖE-----LVKÖ 1597
Oy 1948 GAPQVMOQIINGQPVSTAVSAPNTVSSIPGÖKSLSATSTNSIGSASÖPPRQÖGVK 2007
Db 1598 YÖHQÖATELRL-----QAHMÖMEKÖRQÖDQÖ 1625
Oy 2008 L--TMAÖLT-ÖLTOGHG--NOGLTVVIOGÖGÖTQLOLIPÖV--TVLPFGQÖQIÖQA 2061
Db 1626 LÖEIKRLRLÖLAQÖGNENÖG---EVEÖQFKREKLDKPEVREPLLISNEYALÖKA- 1681
Oy 2062 MPNGVÖRFLPTPLATTAST-----TTTTVSTAAGTÖGÖRÖSKUSPÖMÖVHQ 2112
Db 1682 --NNRLKILLEVKTAAVEETIGRHVLGIIDRSKSSÖSASLIMRSFAEASVKSCHE 1739
Oy 2113 DKTLR-----PÖSSVGV-----AKAÖPQÖASAPQÖTQÖSPÖPE-----2153
Db 1740 EHTRYTDESIPSTSGDMFRNDINMWSKVTEGETELSQRLVNSGFAGTIDDPNEELMN 1799
Oy 2154 VÖTÖPEVÖTÖTVSSHVPSEAOPTHAÖSSKPÖVAÖS-----QÖPSNVGÖSPVRVQÖPS 2208
Db 1800 ISSRLÖAENKLELAEISÖSSÖLEHAKYVÖTLMRSFPKÖKÖATESLQÖELRRLHE 1859
Oy 2209 ÖTRIRPSTPSQÖPQ-----QÖVÖVÖTT-----SÖPIPTÖPHTSLOIPSQ 2249
Db 1860 ESRAÖRLAVELSKAEGVIDYADEKTEFLERQIOEÖTIDILDRLELLCASRLQELAE 1919
Oy 2250 GÖPÖSÖPÖVÖSSÖTLLSSÖQILÖNVSVSSPRQÖLQÖPÖQVYAVLPÖLQÖVÖVLSÖI 2309
Db 1920 -----QÖÖIÖERRELLSR--ÖKEAMKAÖGP--VEÖ-----ÖLÖDETEKIME 1958
Oy 2310 ÖSÖVVAÖIÖAÖSÖGVQÖYÖK-LÖLPÖIÖOÖSSAVÖTHQÖIÖNY-----VTVQASVÖBÖLÖR 2364
Db 1959 KLEVÖCÖÖKAEKVNDLÖKÖVKALEIVÖEVSARFIEÖKNTNELMDLÖKÖNÖALEKÖLEK 2018
Oy 2365 VÖÖLDRÖÖKÖKKÖÖÖETIKREHTLQASNÖSEIÖK--QVVMKNA-----VIEHLKÖK-- 2416
Db 2019 MKRFLDEQÖLIDREHENDVQÖEIQKLEQÖLVVPRPÖPISÖHÖTREVÖQLANHLEKÖTDK 2078
Oy 2417 -SMTAERENÖRMI--VCNÖVKYIILDKIDKEKÖAKKREESYÖK-----2463
Db 2079 CSÖLLLSKÖLÖRDIOERNETEKLFEFRVRELEÖALLDRKHFAGVAEPKELSEVÖLÖA 2138

```

